SUPPORTING HEALTHY DYADIC HUMAN RELATIONSHIPS WITH POWER DIFFERENTIALS USING ROBOTS

A Dissertation
Presented to
The Academic Faculty

by

Michael Pettinati

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Interactive Computing

Georgia Institute of Technology
May 2020

SUPPORTING HEALTHY DYADIC HUMAN RELATIONSHIPS WITH POWER DIFFERENTIALS USING ROBOTS

Approved by:

Dr. Ronald C. Arkin, Advisor
School of Interactive Computing
Georgia Institute of Technology

Dr. Rosa Arriaga
School of Interactive Computing
Georgia Institute of Technology

Dr. Sonia Chernova
School of Interactive Computing
Georgia Institute of Technology

Dr. Ayanna Howard
School of School of Electrical and
Computer Engineering
Georgia Institute of Technology

Dr. Mark Riedl
School of Interactive Computing
Georgia Institute of Technology

Date Approved: April 10, 2020

To Mom and Stephen

You couldn't have anticipated the ride but stayed with me the whole way. Thank you.

ACKNOWLEDGEMENTS

First and foremost, I want to thank my advisor Dr. Ronald C. Arkin. Without his guidance, this thesis would not have been possible. It was work I began with Dr. Arkin that prompted me to ask the research question that drove this dissertation. As I carried out the work required to answer this question, he provided instrumental direction and discerning feedback. His support was critical in the completion of this work.

I would also like to thank my committee members: Dr. Rosa Arriaga, Dr. Sonia Chernova, Dr. Ayanna Howard, and Dr. Mark Riedl. They provided invaluable support and insights as I designed and carried out my human-subject studies. Thank you so much for your time, encouragement, and care.

This thesis would not have been possible without the support of all of my family and friends. I'd like to highlight a couple of individuals in particular. Thank you to my lab mates: Shu Jiang, Matthew O'Brien, Jaeeun Shim, and Lakshmi Velayudhan. Matt, Jaeeun and Lakshmi you all dedicated a considerable amount of your time and energy to help me complete my human-subject studies. You fulfilled roles that I could not fulfill myself. I am truly grateful for all of your help, and I hope to work with you in some capacity in the future. You all share a strong work ethic, intelligence and good humor. Shu, although you never participated in my HRI studies, you were the one who taught me MissionLab. It's hard to overstate how meaningful a contribution that has been to my work. In addition, I have to say thank you to Matt and Shu for the enjoyable lunches during the week that always helped me to decompress.

I'd like to recognize my brother, Stephen Pettinati, and my mother, Susan Gobron. You have both listened to a lifetime worth of my complaining. I can only hope that in the future I have the opportunity to provide you with the same level of assuredness and comfort that you have provided me. In addition, I have to recognize my aunt, Pamela Pettinat. Your tenacity has always inspired how I've approached my education and any professional pursuit. Thank you for all you have taught me over the years, the advice you've given me, and the deep belief you have always had in me. I would be remiss not to mention my friends: Jake Christensen, Joshua Rosenthal, Paige Crawford and Erin Cheak, who have also provided immeasurable support to my wellbeing. There have been countless times when you all have helped restore my sanity when I thought it was lost and then derived pleasure depleting it again. Thank you.

Finally, I have to mention my father, Joseph Richard Pettinati. You didn't get to see much of this part of my life. You were a major part of it though. It was the hopes you had for my future and your never wavering belief in me that sustained me when I had trouble believing in myself. I only wish we could celebrate it together when it's over.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

LIST OF ALGORITHMS

SUMMARY

Conflict is a natural part of ever-evolving human-human relationships. The way in which conflicts are handled can result in relationship growth, dissatisfaction (for one or both parties), or relationship dissolution. Hierarchical relationships often handle conflict in ways that result in the lower-power member being dissatisfied with the disagreement's outcome. This thesis explores how a robot, a piece of technology that can be persistently present during hierarchical human-human interactions, can help to support healthier conflict processes in these relationships when humans cannot.

In order to help, the robot must be able to identify problematic relationship states in the interactions. This work introduces a novel and general computational model to identify six problematic relationship states seen in hierarchical relationships. It couples this model with original robotic behaviors and an action-selection mechanism that allows a robot to successfully intervene to support the amelioration of these problematic states. This computational architecture is operationalized and partially implemented building upon work from human mediation, conflict theory, and occupational therapy.

An intervening autonomous robot that is responsive to problematic relationship states using this computational architecture is compared to an unengaged robot (a robot that is not responsive to the human-human conversation) using a series of human-robot interaction (HRI) studies. The results of these studies show the potential for robots to make positive differences in human-human communication and suggests future research directions for prosocial technologies.

CHAPTER 1

INTRODUCTION

A flushed-faced parent and dejected adolescent head off in opposite directions at the conclusion of an extended and emotionally charged interaction. The exchange concluded suddenly with the parent issuing a definitive, “Because I said so”, in response to the adolescent asking why he could not go out with his friends on a school night. The two parties had presented their views repeatedly over the course of several minutes; each presented his side while ignoring the other’s case. The parent, feeling disrespected, used his authority over the adolescent to end what seemed to be an unproductive discussion.

The resolution of this overt conflict does little to improve the parent-adolescent relationship. The relationship is still in a state of dissatisfaction. The parent sees his child carelessly making irresponsible decisions that could negatively affect his future. The adolescent feels that the parent is overbearing and unwilling to let him assert his independence. Both sides feel weak; each was unable to garner recognition from the other. If the parent-adolescent relationship is to return to a state of satisfaction, they need to have healthier interactions going forward where they clearly express their own feelings while acknowledging each other’s point of view.

In dyadic relationships, where a power difference exists, the ability of the more powerful member and/or the desire of the weaker party to avoid overt conflict often leads to unhealthy conflict practices that leave the relationship in an unsatisfactory state. People can recruit professional mediators or therapists to help them improve negative conflict interactions; however, professionals are unlikely to be recruited for day-to-day matters. An agent that facilitates the use of healthier interaction styles would be invaluable in these moments for these relationships. This thesis explores how a robot can fulfill this role and help human dyads have more positive and open interactions.

A robotic agent that can support a healthy relationship between two people would provide immeasurable social benefits. As robots enter homes, businesses, and medical facilities, not only as tools for domestic services, but also as companions, they have a unique opportunity to support healthy dyadic human relationships. The persistent presence of robots would allow people to improve relationships when unhealthy interaction patterns first arise, before relationships have the chance to worsen, and before professional help may be required. Those closest to the dyad members may be present and able to intervene in order to improve unhealthy patterns of interaction; however, these interventions often do not help the dyadic relationship. For example, nonprofessional interventions can disrupt relationship development (e.g. Beisecker, 1989), complicate the conflict process, and push toward conflict resolution that is not in the interest of the dyadic relationship (e.g. Vuchinich et al., 1988; Taylor, 2002).

If a robotic agent can model a dyadic human relationship and recognize unhealthy conflict, then the agent can limit its presence by only intervening in situations that are potentially destructive. This would allow dyads to take a healthy approach to conflict management to develop and restore their relationships when conflict arises. When a robotic agent supports the dyad, it can take precautions to avoid complicating the interaction. The capabilities, the goals, and role of the agent can be made clear to those involved. The parties can be assured that the agent is unbiased in its role and is present to support both parties. Dyad members need not have a complicated history with the robot. They likely have complicated histories with the people around them. A dyad member's history with a nearby person could make that person a poor choice for intervening in a negative conflict. A neutral robot could avoid some shortcomings that are inherent with a third person trying to support a healthier interaction.

The notion of a robot supporting healthier interactions between two people is an area that has recently gained interest among researchers (e.g. Hoffman et al., 2015). The research up until this point, however, has depended on identifying simplistic cues, such as sustained shouting (Hoffman et al., 2015) and scripted insults (Jung et al., 2015), to detect conflict in a relationship. These cues are not sufficient to determine when and how a robot should intervene when considering the long-term health of a relationship (Noce, 2010). Further, the prescriptive interventions used by these robots do not fit relationship-focused interventions strategies used by human practitioners (e.g. Bush & Folger, 2010a). Up until this point, to the best of our knowledge, researchers have not proposed a real-time system that focuses on providing support for healthier dyadic human relationships. A relationship-focused tool may allow for the natural development of relationships that use healthy interaction strategies, while supporting the use of healthier strategies in times when there is relationship strain.

The central thesis is that a framework can be developed and embodied in an autonomous robotic system that is able to support communication in hierarchical human-human relationships.

The subsequent chapters of this dissertation:

- Develop a general computational model for an autonomous robot to identify problems in hierarchical human-human relationships; this model is operationalized using relationship-focused mediation literature.

- Develop a general action-selection mechanism for an autonomous robot to intervene to support hierarchical human-human relationships. This architecture is operationalized and implemented using relationship-focused mediation theory.

- Articulate and follow an extensively researched and justified methodology to construct and to test the computational model and action-selection mechanism in human-robotic interaction studies.

- Present results showing how these mechanisms can work together to meaningful change how one dyad member perceives the other during conflict. Additional results suggest the agent can support positive change in hierarchical dyadic relationships.

1.1 Research Questions

1.1.1 Primary Research Question

How can social robots be used to support positive change in relationships with power differentials that are experiencing negative conflict?

- This question motivates all of the work that has been completed and contained in this dissertation. In relationships with salient power differentials, there are often unhealthy communication or interaction patterns that lead to relationship dissatisfaction. Unhealthy conflict can be detrimental to both members of the dyad. The dynamics of negative conflict are covered in Chapter 2. A robot may be able to model a dyadic human relationship and understand when there are unhealthy patterns of exchange. When unhealthy communication patterns emerge, the robot can use subtle cues to support healthier interactions. A computational model and action-selection mechanism to support relationships is introduced in Chapter 3. Chapter 4 operationalizes and justifies the implementation of this computational architecture. It describes how states of negative conflict are identified, what constitutes positive change, and how a robot can enact this change. Subsequent chapters describe studies to validate the architecture.

1.1.2 Subsidiary Questions

How is an unengaged robot, that has been introduced as something that is meant to support communication within the relationship, perceived by the dyad members and how do interventions change this perception?

- An unengaged robot is an agent that moves in subtle ways but is not actively responding to the interaction at hand. This is juxtaposed to an intervening robot; its behaviors depend on how the dyad is conversing. It is important for both dyad members to see the robotic agent as a neutral tool that is not on a side. Further, it is important for the robot not to feel disruptive. The mere presence of an agent that is said to be modeling or supporting the relationship's communication may change members' behavior. The higher-power relationship member may perceive or react differently to a robot than the lower-power relationship member. As discussed in Chapter 2, a person's understanding of an agent's capabilities and purpose can influence perceptions of the agent. The experiments discussed in Chapters 5 and 6 examine perceptions of an unengaged agent introduced as something that may support communication versus that of an intervening robot from the perspectives of higher- and lower-power dyad members.

How can a social robotic agent represent a dyad's relationship state, problematic or otherwise, and decide when to intervene in the relationship?

- In order for a robotic agent to know when it should act in order to help to ensure a satisfactory relationship between dyad members, it needs to have an understanding when the relationship is not satisfactory or is in a destructive state. A computational model to classify the relationship state is presented in Chapter 3. The model is operationalized and implementation details are given in Chapter 4. The experiments to validate this computational model are covered in Chapters 5 and 6.

What channels of communication should the agent use when supporting the dyad, how overt should these communications be in order to avoid upsetting the relationship, and how should the agent choose between interventions?

- It is critical that the robotic agent allows the dyad members to interact in a natural way. That is, the robotic agent should not intervene in extremely overt ways that disrupt the dyad's communication and permit them to focus on one another, unless absolutely necessary. This dissertation will try to understand how to effectively and subtly help dyad members recognize and repair unhealthy patterns of communication. The literature underlying intervention strategies is introduced in Chapter 2. An action-selection mechanism is introduced in Chapter 3. The implementation details of interventions based on the literature are provided and justified in Chapter 4. These interventions were explored as part of a pilot study and are validated as part of the complete study covered in Chapters 6.

What issues does the power dynamic in the relationship present for the robotic agent when trying to provide a conflict process that is viewed as fair and equitable by both participants?

- In certain contexts, it may be the obligation of the more powerful dyad member to foster mutual understanding and help the relationship to heal. If the dyad member with more power is failing in this role, it may seem reasonable to take the weaker member's side in the conflict. It is important, however, not to shame or alienate the dyad member with more power. The studies covered in Chapters 5 and 6 put participants in higher- and lower-power roles to understand how the robot is perceived, how the other individual in the relationship is perceived, and how the agent's interventions may alter the behavior of individuals in roles with differing amounts of assigned power.

1.2 Thesis Contributions

This dissertation is about understanding how a robotic agent can aid in the attainment of more positive and open human relationships where a power difference exists. In order to successfully aid these relationships, while not being unnecessarily disruptive to these relationships, the agent must have an understanding of what is healthy and unhealthy behavior in different contexts and relationships, what the current state of a relationship is, and how to act to bring about or preserve relationship health given the current relationship state. Accordingly, this dissertation will make the following contributions to the robotics community.

- This dissertation presents a general computational framework to model the health of a dyadic human interaction with a power difference.

- A computational model that assesses relationship health has been implemented and tested. It was coupled with an action-selection mechanism to effectively attain a more positive and open relationship in the experiment’s context.
- Guidelines for how this general framework can be applied or extended are given as part of the dissertation.

1.3 Thesis Overview

In relationships where a power differential exists, negative conflict management methods can often be the preferred methods of conflict management. This introductory chapter suggests that a robotic agent may be able to help ensure more positive, open and constructive conflict interactions in human dyads. Chapter 2 more closely examines conflict management in dyads and introduces relationship-focused human mediation techniques. It also discusses how technology has already been applied to enhancing human relationships. The related work has relevance to all four subsidiary questions.

Chapter 3 proposes a computational model and action-selection mechanism that allow for a robot to understand and appropriately intervene in relationships where the mutual satisfaction is threatened. This chapter addresses the second and third subsidiary questions. Chapter 4 operationalizes the computational architecture and provides implementation details, which are relevant to the second and third subsidiary questions.

Chapter 5 provides the methodology and results of a human-robot interaction (HRI) study. This study was concerned with understanding how people with different levels of power perceived a robot meant to help with communication. It served as a baseline for a second experiment. Further, it provided data to create the computational model. This chapter addresses subsidiary questions one, two and four. Chapter 6 discusses insights gleaned from a pilot study used to test the relationship focused interventions as well as a study that builds upon the first to test these interventions. This work helps to answer all four subsidiary questions. The final chapter summarizes insights gleaned from this research and future directions for the research.

CHAPTER 2

RELATED WORK

Two people are in a satisfying relationship when it meets the needs of both members (Rusbult et al., 1998). As time passes, the needs and responsibilities of one or both members may change. These changes redefine the relationship; the requirements each member can expect it to fulfill are different. Conflict or disturbances in a relationship can be a symptom of ongoing or impending renegotiation (Riesch, Jackson, & Chanchong, 2003; Walton, 1969). Conflict is not inherently bad, and its advent does not mean that the relationship is changing for the worse. Studies have found that the result of conflict is often an enhanced mutual understanding between the parties (Kantek & Gezer, 2009; Rahim et al., 2000). A conflict's outcome may be an altered but stronger connection. For example, see Figure 2.1.

Figure 2.1. A hypothetical parent-child relationship as it evolves over time. (a) The parent and child are mutually satisfied with a purely dependent relationship when the child is young. (b) The relationship becomes strained as the child becomes increasingly independent. (c) Mutual satisfaction returns when the parent and child agree upon the needs and wants each have for the other person given how their lives have changed.

When a girl is young, her and her mother may have very positive and fulfilling relationship with the mother in the caregiver role and the daughter in a purely dependent role. However, as the daughter begins to grow and asserts her independence, this relationship can become strained, as the mother's protective role seems to encumber the daughter's growth or the daughter's decision-making seems to contradict what the mother sees as appropriate. As the mother begins to recognize the daughter's independence, and the daughter begins to see the mother as a resource for advice and care, the two can have a positive and fulfilling relationship with a less dependent and friendlier composition.

The goal of a relationship-focused robotic agent is not trying to prevent conflict all together; it is to try to support a healthy conflict process. That is, it is important for the robot to enable dyads to recognize when communication is hindered in some way and help parties to openly acknowledge their thoughts and feelings to one another to allow for the parties to work through transitions together.

There are many "types" of human mediators that help parties deal with conflict, and mediators are often classified based on their relationships with the parties in conflict (Moore, 2014). This thesis focuses on neutral (independent) mediation strategies. There are reasons to make the robot a neutral party that are discussed in Section 2.3 below; additionally, relationship-focused mediators are often most effective when neutral (Moore, 2014). Table 2.1 (on the next page) summarizes how independent mediators can be further classified and how these classifications influenced our approach.

Table 2.1: Different ways of classifying a mediator's role and involvement during a mediation session. This table assumes an independent mediator (a mediator that does not have a relationship or power over either party).

Summary of Different Styles of Human Mediation		
Characteristics of Mediation Session	Range of Possibilities	Robotic Agent's Approach
Mediator's Focus For The Session	• Support or Enhance Relationship • Organize or Enhance Process • Provide Advice or Assessment on Issues • (Moore, 2014)	• Relationship-focused interventions • Keep the agent neutral and minimize instructiveness (see Section 2.3)
Mediator's Degree of Control Over Process	• Low to High • Can be peripheral to the parties (merely motivate dispute resolution) • Can entirely plan the structure of an interaction or even act as a go-between when parties won't speak face-to-face. • (Elangovan, 1995; Lewicki et al., 1992; Nugent & Broedling, 2002)	• Low process control • Robot wants to minimize instructiveness (see Section 2.3). • Parties have better understanding of what they need to discuss and how they need to discuss it (Bush & Folger, 2010a).
Mediator's Degree of Control Over Decision	• Low to High • The third-party has no power over the dyad to force a decision (someone unrelated to either party). • The third-party can force a decision on the parties (e.g. an arbitrator). • (Elangovan, 1995; Lewicki et al., 1992; Nugent & Broedling, 2002)	• No decision control • The agent does not provide any assessment of the substantive material discussed. • The agent has no ability to enforce decisions made by parties.

The following subsection discusses how transformative mediators (a type of relationship-focused neutral mediators) define destructive conflict and provides examples of why certain relationships where power differences exist fall into unhealthy conflict interactions. The second subsection gives background on how transformative mediators help to support healthier interactions between two people. The chapter concludes with an overview of research examining how technology has been used to encourage healthier human-human interactions. The related work provides insight into all four of the subsidiary questions introduced above.

Professional human mediators, in order to help assuage destructive conflict in dyadic human relationships, must be able to identify such conflict. The models and intuition garnered by professionals during their practice can be used to develop a grounded model that a robotic agent can use to detect or classify when dyadic interactions are destructive (see Chapters 3 and 4). An understanding of how mediators define healthy and unhealthy conflict is a critical piece of answering the second subsidiary question. The second subsidiary question asks how a robot model a human-human relationship to recognize strain.

Though a robot and a human mediator have clear differences in their capabilities, the role a human mediator plays in a strained dyadic human relationship can help to inform how a robotic agent is introduced and behaves in a relationship that is in a state of conflict. As discussed below, there are principles that human mediators follow that help them improve unhealthy human interactions. The way in which human mediators follow these principles varies. A robotic agent's implementation of these principles will vary from human mediators because of its more limited capabilities, but they can still guide it.

Literature that discusses how technology has been used to aid human relationships can show successful implementations of certain principles and reveal where improvement is needed. This background is essential for answering subsidiary questions one and three. The first question asks how technology to help human-human communication is perceived. The third subsidiary question asks how a robot should intervene in order to ameliorate strained human-human interactions.

Finally, the way in which human mediators treat power differences during a destructive conflict can help to answer the fourth subsidiary question. This fourth question asks how power differences in the relationship might influence the robot's approach to supporting strained interactions. Mediators' responses to power differences during conflict has been developed and proven through practice; it is important to draw on their expertise and proven success when considering how to handle the power differences in the contexts we are examining.

2.1 Defining Destructive Conflict

Weakness on both sides is, as we know, the motto of all quarrels.

-Voltaire

Conflict is likely to arise in relationships that are in states of dissatisfaction. Conflict can be defined in terms of its causes (Lewicki et al., 1992). Rahim and Bonoma (1979) define conflict in terms of three potential causes. First, conflict may arise because the "behavioral preferences" of one member of the relationship are perceived to be "incompatible" with those of the other member in the relationship. Second, there could be a dispute over scarce or seemingly scarce resources desired by each party. Finally, one member's "values or attitudes" could be exclusive or appear to be exclusive to the "values or attitudes" of the other party. Conflict stems from a relationship member's needs not being met by the other relationship member, i.e., one relationship member is behaving in a way that is undesirable to the other, taking resources that are required by the other, or excluding the other. Conflict is the process by which relationship members recognize and address what is causing dissatisfaction in their relationship.

A healthy conflict process, one which can help to restore a satisfying relationship, requires each party in a relationship to care highly about the needs and wants of the other as well as the needs and wants of the self (Rahim et al., 2000; Lewicki et al., 1992). If both members of a relationship want to meet each other's needs, then it gives the relationship the best chance of returning to state where the needs of both members are actually met. As discussed below, however, it is often the case that healthy conflict processes are avoided and more covert and destructive practices are chosen instead.

These unhealthy types of practices are especially common in relationships where a power differential exists (e.g. Brockman et al., 2010; Riesch, Gray, Hoeffs, Keenan, Ertl, & Mathison, 2003; Riesch, Jackson, & Chanchong, 2003; Tickle-Degnen et al., 2011). Dyad members may elect to avoid or curtail overt conflict by lowering the importance they place on their own needs (e.g. Brockman et al., 2010), particularly if they are the weaker parties in their relationships. Relationship members who are the stronger parties in their relationships may place a lower importance on the other's needs and try to force an end on an overt conflict (e.g. Riesch, Jackson, & Chanchong, 2003; Riesch, Gray, Hoeffs, Keenan, Ertl, & Mathison, 2003). There are relationships where the member with the more power may not even recognize the needs of a party with less power (e.g. Tickle-Degnen et al., 2011).

It is problematic whenever a healthy and overt conflict process that addresses the issues in the relationship and where both members' needs are heard does not occur; the relationship is allowed to subsist in a state of dissatisfaction. Below, there are examples of each of these types of the unhealthy or destructive conflict strategies in relationships where power differentials exist. The examples from three different relationship types reveal a pattern that defines destructive conflict, a pattern that is seen across hierarchical human relationships. See Figure 2.2.

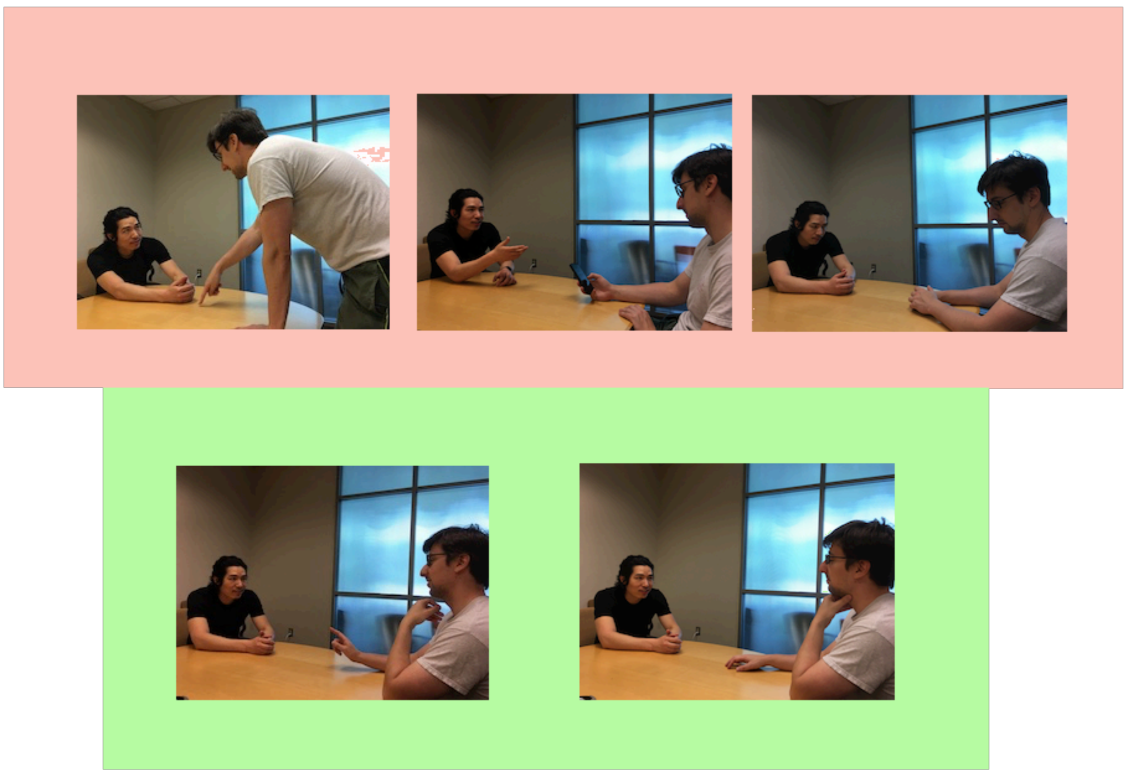


Figure 2.2: Dominant and avoidant approaches do not address the needs and wants of the lower-power member in the relationship. **It is only through direct engagement that both sides understand one another and work towards a positive-sum solution.** The pictures on the top row of the figure show dominant and avoidant (undesirable) approaches. The higher-power dyad member in white is commanding the lower-power dyad member in black to follow instructions in the picture on the left. In the center picture, the higher-power member plays with his phone, ignoring the other. Finally, in the picture on the right, neither dyad member looks at the other. The pictures on the bottom row of the figure show direct engagement. The dyad is looking at each other and considering each other's ideas.

There are sometimes unhealthy interactions in relationships between early-stage Parkinson's patients and their caregivers (Tickle-Degnen et al., 2011). Early-stage Parkinson's patients may have limited expressivity in their faces, bodies, and in their tone of voice, a condition known as an expressive mask (Tickle-Degnen & Lyons, 2004). The lack of expressivity across nonverbal channels (the expressive mask) causes the caregivers of these patients to unfairly attribute negative stereotypes to them; caregivers were found to judge patients with higher masking as apathetic, less competent and extroverted, as well as more depressed and neurotic than those with less masking (Tickle-Degnen & Lyons, 2004; Tickle-Degnen et al., 2011).

When stereotyping the other person in dyadic relationship, one will often over accommodate or under accommodate the other (Williams & Nussbaum, 2001). In the case of over accommodation, one can seem patronizing. In the case of under accommodation, one can seem as if she does not care about the other (Williams & Nussbaum, 2001). The caregiver, viewing the patient through the lens of these negative stereotypes, treats the patient as incapable and/or undesirable. This inappropriate and alienating treatment by the caregiver, based on these stereotypes, is going to make the patient feel ashamed. A person is said to feel shame when "real" fundamental flaws of the self have been revealed, flaws that limit the possibility for positive relationships going forward (Haidt, 2003; Sabini et al., 2001).

Chronically ill patients need their physicians to be empathetic, to help them manage their shame and respond with sensitivity (Zinn, 1993). The caregivers of Parkinson's patients are not fulfilling this need; they are causing the patient shame. Shame in chronically ill patients can cause them to alienate themselves from the caregiver and/or show anger towards the caregiver (Zinn, 1993). When one relationship member responds to the other with hostility, it can cause weakness in the other and further alienate the two (Bush & Folger, 2010a). As seen below, mutual weakness and alienation characterize destructive conflicts.

Conflict is not an uncommon occurrence in relationships between students and their faculty advisors (Adrian-Taylor et al., 2007; Kantek & Gezer, 2009). Further, Adrian-Taylor et al. (2007) found that almost one-third of graduates students who had experienced conflict with an advisor felt that the conflict had not really been resolved. This was compared to just over one-tenth of faculty who said that conflicts with students had not been resolved. This disparity could stem from students not considering their needs highly during conflict resolution. Brockman et al. (2010) says that graduate students often choose approaches that do not consider their needs highly when approaching conflict resolution. Students recognize that they are less powerful than their faculty advisors and ongoing conflict with their faculty advisors can have profound negative consequences for them (Brockman et al., 2010).

Adrian-Taylor et al. (2007) found additional support for unhealthy conflict in student-faculty relationships when investigating the nature of conflict between graduate students and their advisors. Faculty and students do not agree on many of the sources of conflict in their relationship (Adrian-Taylor et al., 2007). Faculty advisors often see the students not fulfilling their needs in the relationship (e.g. the students do not have enough research experience to be contributing as they should); whereas, students often see faculty as not meeting their needs in the relationship (e.g. the faculty do not have enough time for them) (Adrian-Taylor et al., 2007). The fact that each side sees different sources for their conflicts shows a mutual alienation. Adrian-Taylor et al. (2007) hypothesizes that this mutual blame could stem from a "self-serving attribution bias" to protect the self. If true, this shows that the conflict weakens both parties sense of self. It will be important to identify mutual weakness and alienation when devising a model for the relationship (addressing the second subsidiary question in Chapter 1).

There is the common stereotype that adolescents and parents are involved in constant and heated conflict throughout the period of adolescence. In the past, researchers have claimed that "the storm" of adolescence was largely made up; many youths reported satisfying relationships with their parents (Montemayor, 1983; Steinberg, 2000, 2001). It appears, however, that the stereotype of parents and adolescents being involved in frequent conflict is not a total fallacy; the conflicts just take place about common things in day-to-day life (Montemayor, 1983; Riesch, Gray, Hoeffs, Keenan, Ertl, & Mathison, 2003; Riesch, Jackson, & Chanchong, 2003; Steinberg, 2000, 2001).

Parents and adolescents will often not employ healthy conflict strategies when managing these day-to-day conflicts (Riesch, Gray, Hoeffs, Keenan, Ertl, & Mathison, 2003; Riesch, Jackson, & Chanchong, 2003). Parents, who are in a position of power over their adolescents, were found to often use a dominant approach to managing conflict, commanding adolescents to follow what they said (Riesch, Gray, Hoeffs, Keenan, Ertl, & Mathison, 2003; Riesch, Jackson, & Chanchong, 2003). Adolescents, recognizing the adult's power, would often oblige parents to avoid conflict, even if they did not agree with what they were doing (Riesch, Gray, Hoeffs, Keenan, Ertl, & Mathison, 2003; Riesch, Jackson, & Chanchong, 2003).

The parents' and the adolescents' behaviors reveal that they are alienated from one another and weak (Bush & Folger, 2010b). The alienation is clear in that they do not understand one another's needs. They show weakness in their inability to work through a situation that they cannot control; they are both simply giving up on the process that might lead to a more satisfactory conclusion (Bush & Folger, 2010b).

These different examples have shown that conflict weakens both dyad members, and alienates the dyad members from one another. This pattern is something that has been identified by transformative mediators as characterizing negative conflict interactions (Antes, 2010) and is the pattern that needs to be corrected in order to have more positive and constructive interactions moving forward (Bush & Folger, 2010a). This is the pattern that this thesis will focus on identifying when modeling the dyadic human relationship and answering the second subsidiary question from Chapter 1. This is the focus of sections 3.1 and 4.1 below.

Practitioners of transformative mediation, a relationship-focused style of mediation, believe that "conflict, along with whatever else it does, affects people's experience of both self and other" (Bush & Folger, 2010a). People, regardless of their position of power within the relationship, have a sense of "weakness" or "incapacity" during conflict; they have lost their control over the interaction (each person is challenging the other) (Bush & Folger, 2010a). This lack of control threatens relationship members' view of themselves as competent enough "to handle life's challenges" (Bush & Folger, 2010a). This threat to the self causes the people involved in conflict to alienate themselves from the other, to become self-absorbed (Bush & Folger, 2010a). A person embroiled in a negative conflict interaction feels the need to protect the self and becomes more “suspicious”, “hostile”, and “closed” toward the other (Bush & Folger, 2010a). During conflict, the other is the object threatening the self. The self feels unable to provide for the needs or wants of the other and the needs and the wants of the self.

The feelings of weakness and the move toward self-absorption naturally reinforce one another (Bush & Folger, 2010a). As one feels weaker, the individual becomes more hostile toward and closed off from the object that is causing that feeling, namely, the other (Bush & Folger, 2010a). As an individual becomes more closed off from and hostile toward the other, she will treat the other inappropriately, which will lead to increasing feelings of weakness in the other, which will lead to greater hostility in the other, and so on (Bush & Folger, 2010a). The spiral continues until the parties are totally alienated from one another. This destructive cycle mirrors the unhealthy pattern recognized in the relationships above. See Figure 2.3.

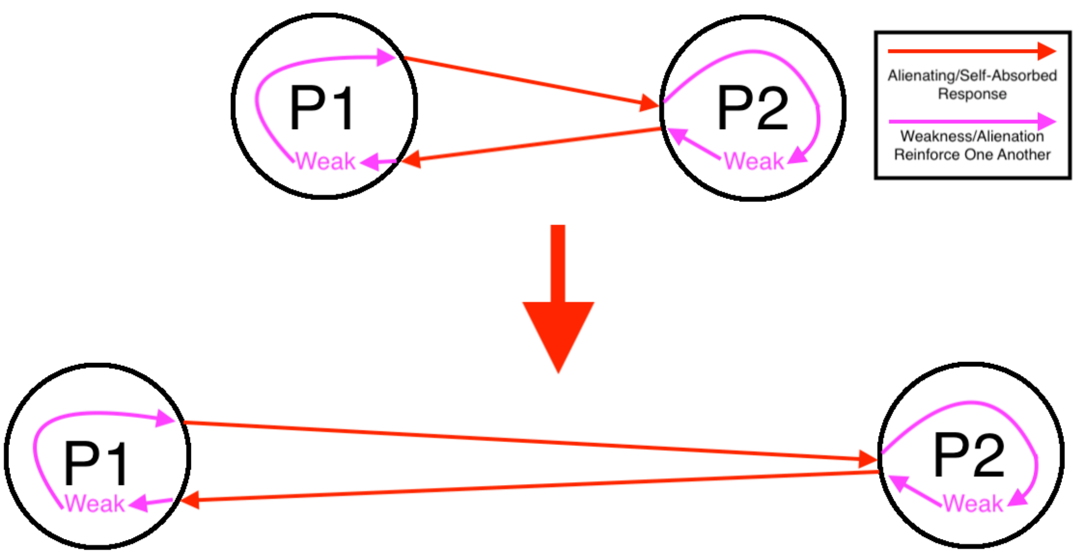


Figure 2.3: When an individual (P1) is weak and does not know how to appropriately respond to another (P2), she may begin to create distance between herself and that other person. If this alienating response is met with a lack of empathy or understanding and responded to with an alienating response in kind, the two enter a vicious cycle where the push each other away until they become completely alienated.

This cycle can be reversed such that the dyad members strengthen each other; a transformative mediator's goal is to reverse the cycle. A critical piece of this thesis will be to understand how to approach reversing this cycle. This will be the answer to the third subsidiary question from Chapter 1. See sections 2.2, 3.2, and 4.2 below.

2.2 Relationship-focused Transformative Mediation

Transformative mediators are responsible for supporting empowerment and recognition shifts in dyadic relationships that are in a state of negative conflict (Bush & Folger, 2010a; Della Noce et al., 2003). Supporting moves of empowerment for the dyad members are particularly important because recognition will generally follow empowerment; as a person's self becomes stronger, she will be more open and receptive of the other (Bush, 2010). This positive and constructive cycle is visualized in Figure 2.4.

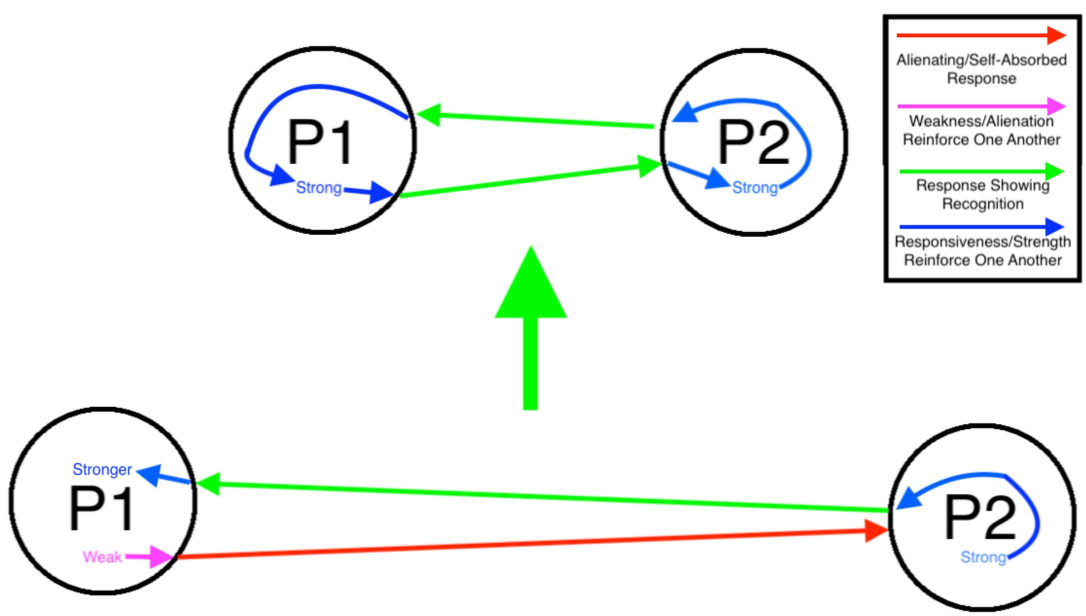


Figure 2.4: When an individual (P1) is weak and tries to push away from the other person, and that alienating response is met with understanding and empathy by the other (P2), the initial individual become stronger. A positive, constructive and mutually empowering cycle begins that brings the two parties closer together.

A major factor in enabling an interaction shift is the framing of the interaction and the mediator's role. The mediator needs to make clear that she trusts the parties and their abilities to make decisions (Bush & Folger, 2010a; Bush & Pope, 2002). It is important that the dyad members understand that the interaction is intended to be a conversation between the two of them and that the mediator is simply there to assist the two of them (Della Noce et al., 2003). The tools the mediator uses are always intended to "support" the dyad's decision-making, never to "force" the dyad in a certain direction (Bush & Folger, 2010a; Bush & Pope, 2002; Della Noce et al., 2003).

A transformative mediator provides this support through "reflection", "summarization", "checking in" and "questioning" (Bush & Folger, 2010b; Bush & Pope, 2002). In reflection and summarization, the mediator will echo what one or both of the dyad members have stated during the conversation; the mediator maintains the content as well as the "emotional tone" of the parties (Bush & Folger, 2010b; Bush & Pope, 2002; Della Noce et al., 2003). The mediator uses reflection and summarization to clarify and amplify the conversation (Bush & Folger, 2010b; Bush & Pope, 2002). These tactics give the dyad members the opportunity to hear what they have been discussing and how they have been discussing it, where the similarities and differences between them lie, and consider, given this information, where they want to go from here (Bush & Folger, 2010b; Bush & Pope, 2002; Della Noce et al., 2003). Our agent will draw inspiration from these tactics to reverse a negative conflict cycle (see sections 3.2 and 4.2).

Reflections and summaries should always be done tentatively; that is, mediators should check with the dyad members to ensure that the reflection/summary made was accurate and allow for one or both of the dyad members to clarify what they are saying

(Bush & Folger, 2010b; Bush & Pope, 2002; Della Noce et al., 2003). The dyad members need to have the opportunity to consider how they were perceived, consider what was taken away from their statement, and restate or elaborate on what they previously presented to better present their feelings and position (Bush & Folger, 2010b; Bush & Pope, 2002; Della Noce et al., 2003).

The fact that the mediator lets the parties make their own decisions and carryout their own process supports the empowerment of the dyad members (Bush & Folger, 2010b; Bush & Pope, 2002; Della Noce et al., 2003). If the mediator were to push the dyad members in a certain direction or make the decisions on behalf of the dyad, then it brings into question the dyad members' competence and weakens them further (Bush & Folger, 2010b; Bush & Pope, 2002; Della Noce et al., 2003). When the mediator lets the dyad make their own decisions, this shows an implicit trust in the dyad members. The reflections and summaries given by the mediator help the dyad members to hear themselves and express themselves clearly (Bush & Folger, 2010b; Bush & Pope, 2002). The ability to express oneself clearly and as intended is empowering; therefore, providing the information that enables dyad members to express themselves more confidently is supporting an empowerment shift (Bush & Folger, 2010b; Bush & Pope, 2002).

2.3 Technology and Human Relationships

Robotics researchers have examined a small portion of the problematic dynamics that arise in human-human relationships and human groups and have provided preliminary insights into how robots as well as other feedback systems might be able to help address these issues. Jung (2016) and Sonalkar, Jung, and Mabogunje (2011) show how the same types of socio-emotional dynamics that are often detrimental to marriages in the long term also hurt the long-term success of student engineering teams. Expressions of hostility and the disproportionate expression of negative affect (compared to positive affect) during thin slices of conflict at the beginning of semester-long student projects were predictive of poorer final projects from student teams. Jung (2016) advised that a robot could model the emotional dynamics of interactions and help groups balance positive affect and negative affect as well as repair hostility to enhance their long-term outcomes. Robotic interventions to help relationships subsist or have task-based success, however, may differ from trying to support mutual empowerment and recognition to attain a more satisfying relationship.

Jung et al. (2015) examined a robot's ability to repair hostility within a group setting. In this experiment, an experiment confederate directed a tasked-based critique or a personal insult at one of two participants. A robot within the group responded to the confederate with a neutral statement or a repair statement (the repair statement chastised the confederate and requested the group stay positive).

When it comes to supporting feelings of mutual empowerment and recognition, it is critical to encourage open communication and combat suppression. The overt verbal interventions of the robot drew attention to the social norm violation; however, they did not support the participants' abilities to speak openly and acknowledge the emotion.

Shen et al. (2018) showed how a robot (controlled using a Wizard of Oz approach) could be used to help young children identify, work through, and reach resolutions to conflicts. The robot seemed to "interrupt the emotional and reactive momentum" that builds before a conflict, and it provided control over the conflict process allowing for many children to reach substantive agreements. Again, if the argument is purely substantive (e.g. possession over a certain object) this may be desirable; however, if the aim of the conflict process is to reach a more mutual understanding, then limiting emotional expression is not desirable. If the goal is for the technology to support the relationship, it can be problematic it to control the interaction in any way.

Hoffman et al. (2015) presented a peripheral robotic lamp that served as a conversation companion. This lamp tried to discourage yelling during interactions between married couples (Hoffman et al., 2015). The lamp would act afraid (shake) when people's voices exceeded a certain decibel level (Hoffman et al., 2015). The results presented suggest that a peripheral robot can influence a human-human interaction without being a disruption to that interaction (Hoffman et al., 2015).

As stated above, transformative mediators are present during a dyadic interaction to help the dyad members' reverse the negative cycle that defines destructive conflict; a mediator is never there to make decisions on behalf of the dyad members (e.g. Bush & Folger, 2010a). Acting afraid in response to raised voices is making a decision on a dyad member's behalf, however. The decision is that yelling should not be allowed. The lamp is trying to nudge the dyad members to not act in a certain way. Further, research has shown that humans will generally not listen to a robotic agent that commands them to do something (Liu et al., 2008; Roubroeks et al., 2011), and people have negative impressions of other people who are submissive to a robot (Li et al., 2015).

The robotic lamp tried to subtly influence the dyadic human relationships. Hoffman et al. (2015) emphasized the importance of "supplementing" the human relationship, being "peripheral" to the interaction and not directly interacting with the dyad members. Other researchers have stressed helping to regulate emotion without conscious effort (Costa et al., 2018). These researchers changed voice feedback to change self-perception and change emotions during conflict. Participants felt more empowered and calmer. This may eliminate some of the issues with people feeling submissive to the technology; however, again this type of approach is manipulative and not relationship-focused. It is supplanting the dyad's decision-making about how the conflict should be handled and what emotions they are experiencing when dealing with different topics. It is always about supporting decision-making and never supplanting in relationship-focused endeavors (Bush & Folger, 2010b; Bush & Pope, 2002; Della Noce et al., 2003).

DiMicco et al. (2004) attempted to enhance team functioning (by encouraging input from everyone in a group) using a peripheral display that showed the verbal participation rate of all of the members in a group. The presence of this display disrupted the development of trust among group members (DiMicco et al., 2004). DiMicco et al. (2004) believe that the display was disruptive because it publicly displayed "socially sensitive information". Group members could have seen the display as pressuring certain people to talk while silencing others. It is important to make sure the role of the technology and the information it presents is clear.

Tennent et al. (2019) used morphology to make the role of the technology clearer. They had a microphone that participants were meant to speak into. In the condition where the microphone moved to follow the speakers and checked in with the person speaking the least, the groups were more balanced and performed better on the experiment's problem-solving task.   A central tenant of transformative mediation is to ensure that the dyad understands the role of the mediator in the interaction (Bush & Folger, 2010a; Bush & Pope, 2002; Della Noce et al., 2003).

It will be important to detail the capabilities of the agent explicitly when introducing the robot because people tend to anthropomorphize robots (e.g. Duffy, 2003). People may interact with a robotic agent inappropriately because of very simple anthropomorphic cues or very simple behaviors (e.g. Duffy, 2003). The metaphor people use to understand an artificial agent will largely determine how they treat that agent (Sirkin et al., 2015). If dyad members are given an explicit understanding of how an agent is functioning, then they are more likely to communicate with it appropriately.

Some social behaviors that are evoked by technology do appear to be "mindless" or automatic (Nass & Moon, 2000). It is important to understand through testing how the mere presence of a robotic agent, which has been introduced appropriately, with a certain morphology and behaviors, influences an individual's behavior and comfort during a human-human interaction (and how that individual's role affects his/her behavior and comfort).

Beyond the form and behaviors of a robot "mindlessly" influencing individuals' treatment and acceptance of them, there is also the group membership of the robot. There appears to be complex interplay between whether the robot is perceived to be an ingroup or neutral member of a situation involving multiple people, the relationship between the people, and the types of the requests the robot makes when it comes to determining whether or not people will listen to the agent (Sembroski et al., 2017). It is important for individuals to not view the robot as an ally or foe but something trying to understand the situation. It is also important, if the robot is to make a request, it must be framed simply or not require substantial engagement from the person receiving the request.

Transformative mediation allows for dyad members to disagree with the mediator. As mentioned above, mediator reflections and summaries are always done tentatively. A robot that evaluated peoples' work lost their trust when they disagreed with its evaluation (Banh et al., 2015). If the agent is responding tentatively (as if trying to understand the situation at hand), they may not simply disuse the agent when it makes an assessment with which they disagree.

The research into technologies that can help to support human-human relationships is summarized in Tables 2.2 and 2.3 on the next two pages. This work provides insights into all four of our subsidiary questions. The table extrapolates the information covered here and how it can be applied to the system to support positive shifts in human-human relationships or how such a system has to differ.

Table 2.2: Summary of the robotics research related to supporting human-human relationships and how it applies to subsidiary questions one and two from Chapter 1.

Summary of Previous Work as It Related to Subsidiary Questions		
Research Question	Previous Finding Related to Research Questions	Application of Related Work to Research Question
Perception of Robot	• Without clear metaphors, people interact with robots differently (e.g. Sirkin et al., 2015). • Individual differences in how much people will anthropomorphize technology (Fischer, 2011; Nass & Moon, 2000). • People may mistrust or disuse a robot if they disagree with the information it presents or what it is asking of them (e.g. Cormier et al., 2013; Banh et al., 2015). This can be context specific (e.g. Robinette et al., 2016).	• Use a clear and concise metaphor to introduce the robot. • The robot is meant to help support communication. • The robot is not an active participant in the conversation and not knowledgeable about the topic at hand. • Interventions should be tentative (agreeing with the literature from above) to avoid mistrust/disuse of agent.
Representation of Relationship	• Many studies have used a Wizard of Oz approach for intervening in strained relationships (e.g. Jung et al., 2015; Shen et al., 2018) • Interventions for strained relationships have been triggered based on single cues such as raised voice (Hoffman et al., 2015) and set insults (Jung et al., 2015). • Interventions for improved task performance triggered based on objective measures (Zuckerman et al., 2016; Tennent et al., 2019).	• Few relationship-focused intervention systems, if any, consider more than a single cue. • Insights into strained conflict interactions require constellations of cues (Bush, 2010; Noce, 2010). • Draw from mediation literature to devise insights about relationships that may be more robust than current methods (see chapters 3 and 4).

Table 2.2 (cont.): Summary of the robotics research related to supporting human-human relationships and how it applies to subsidiary questions three and four from Chapter 1.

Summary of Previous Work as It Related to Subsidiary Questions		
Research Question	Previous Finding Related to Research Questions	Application of Related Work to Research Question
Relationship-Supporting Interventions	• Technology can help people to recognize the negative emotion during an interaction (Jung et al., 2015; Shen et al., 2018). • Robots can encourage people to act in certain ways (e.g. not shout or to speak more during an interaction) without being a disruption (Hoffman et al., 2015; Tennent et al., 2019). • Technology can alter peoples' perceptions of themselves to help avoid negative affect (Costa et al., 2018). • People will not listen to technology that commands them (Liu et al., 2008; Roubroeks et al., 2011). • People think unfavorable of those who are submissive to technology (Li et al., 2015).	• We want to support open communication. • We want emotions to be expressed and worked through. • We want individuals to consciously engage each other. • Interventions are fundamentally different from those previously proposed. • Important to incorporate mediation literature in designing interventions (see chapters 3 and 4). • Again, there is the notion that interventions should be suggestive/tentative if we are going to have participants consider what the robot has said and appropriately respond.
Ensuring that Interventions Are Perceived to Be Fair and Equitable	• Willingness to listen to and/or act on requests made by the robot is influenced by: • The robot's group membership. • The importance of the request (people listened to innocuous requests). • The authority or relationship to present humans. • (Sembroski et al., 2017).	• The robot should be neutral to avoid triangulation issues. • The robot should treat the parties similarly (not single out or favor either party). • Interventions should be simple/not require something difficult of participants.

2.4 Conclusion

This chapter gave an overview of the research that has been done in conflict theory, human mediation, as well as how technology has been used to support human-human relationships. It began to answer the thesis's second subsidiary question, which asks how a robot could represent a human-human relationship and identify strain. Destructive conflict is defined by mutual alienation and weakness in the dyad members. There are problematic mental states within the individuals that make them weak and want to withdraw from the person with whom they are interacting. They push the other person away. **The robot needs to model problematic mental states that signify weakness (desire to disengage) and hostility (active rejection) within the dyad members to understand when the relationship is strained and successfully intervene**. The model to identify these states is introduced in Chapter 3.

During destructive conflict both members of the dyad are in a weakened state relative to their pre-conflict state (e.g. Bush & Folger, 2010a). As noted above, trying to enforce certain rules, control behavior, or manipulate individuals further weakens the reprimanded/manipulated dyad member. Transformative mediation says **dyad members with different power in the interaction should not be treated differently** (Bush, 2010). The decision-making of both people needs to be supported by a neutral agent (subsidiary question 4 asks how to ensure the conflict process is perceived as fair and equitable by both dyad members).

The dyad needs to be given a clear and concise introduction to the robot; they need to understand its role in the interaction, how it will behave as well as how its behaviors can be interpreted. This introduction is crucial if the robot is to be successful in supporting the empowerment of the dyad members during destructive conflict and helps to ensure the dyad members do not misuse or disuse the agent. Question one asks how a robot to support communication in a human-human relationship is perceived. **The perception of the agent and its use is closely tied to how the robot is introduced.**

When the agent recognizes negative conflict, the agent needs to support empowerment. This does not mean pushing the use of certain interaction strategies. **The robot's interventions should support the sharing of ideas and emotions**. The behaviors the agent can employ to empower the dyad members are further considered in the fourth chapter. These previous results are relevant to subsidiary question three from the first chapter, which asks how a robot should intervene to support human-human relationships.

CHAPTER 3

COMPUTATIONAL MODEL

A robot tasked with supporting positive change in a dyadic human relationship where a power difference exists must first be able to recognize when there is dissatisfaction in the relationship (subsidiary question 2 from Chapter 1); it must then decide how to help to ameliorate that dissatisfaction (subsidiary question 3 from Chapter 1). This chapter introduces a computational model and an action-selection mechanism that work in conjunction to identify and support the betterment of problematic states in dyadic relationships with power differentials. This model was first introduced in Pettinati and Arkin (2018).

The problematic states introduced as part of this model, and the behaviors that support the amelioration of these states, are grounded in literature from Transformative Mediation. Chapter 4 operationalizes the states (section 1) and provides the full implementation details of this computational architecture (sections 2 and 3).

Our robot is playing a purely supportive role in the dyad's relationship; therefore, the interaction between the higher-powered and lower-powered individuals can be characterized as shown in Figure 3.1 (on the following page). The relationship has three actors, A = {H, L, R}, where H is the high-power dyad member, L is the low-power dyad member, and R is the robot. H and L give each other direct attention, while R's presence is acknowledged at the interaction's periphery with passing glances. R is a bystander. R is attentive to both dyad members. It is consistently monitoring what each dyad member is saying and doing.

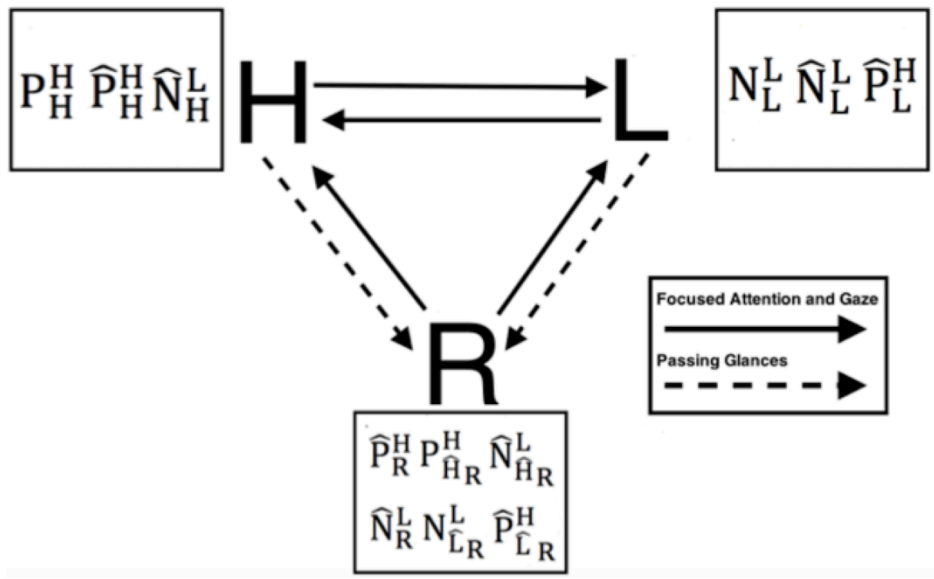


Figure 3.1: The relationship between our three actors and the partial theory of mind considered for each actor

We consider a partial theory of mind of each human actor and the robot's representation of these mental states. The mental states allow us to enumerate problematic relationship states in the following subsection. The mental states of both human actors and the robot's representation are summarized in Table 3.1. As described below, problematic relationship states are determined by percept values that are indicative of active mental states that threaten a positive and open interaction, an interaction that is beneficial for long-term relationship health.

Each mental state component is represented by a scalar value. The magnitude of the scalar value is indicative of the degree to which the mental state component is present in the individual. States are active if the mental states have magnitudes that fall below or above set, acceptable thresholds (see Table 3.2). The percepts related to these mental state components serve as input to a predefined mapping that decides what problematic relationship states are active and what states are inactive. The representation of these percepts is percept specific. Section 3.3 contains a full example of the how the computational architecture is used to identify active problematic relationship states.

In the text, the partial theory of mind of each actor contains components that follow the same notational scheme. The regular sized text is either the letter P or N; this letter indicates whether a positive or negative valence emotion is being represented. A subscript character is indicative of who is evaluating the state (H, L, or R). A superscript character is indicative of who possesses the emotion being represented (H or L). A hat over the letter indicates the agent making the evaluation of the person or emotion is estimating that value.

Table 3.1: Summary of the human actor's mental states representation and the robot's representation of these states.

H's Mental State	$\{P^{H}_{H}, \hat{P}^{H}_{H}, \hat{N}^{L}_{H}\}$
	P^{H}_{H} - H's positive responsiveness toward L of which she is consciously aware (attention and uplift)
	$\hat{P}^{H}_{H}$ - H's estimate of how she is conveying these positive emotions.
	$\hat{N}^{L}_{H}$ - H's estimate of L's negative response toward herself (signs of disengagement and desire to withdraw).
L's Mental State	$\{N^{L}_{L}, \hat{N}^{L}_{L}, \hat{P}^{H}_{L}\}$
	N^{L}_{L} - L's negative thoughts toward the self of which she is consciously aware.
	$\hat{N}^{L}_{L}$ - L's estimate of how she is conveying these negative emotions.
	$\hat{P}^{H}_{L}$ - L's estimate of H's positive emotions toward her.
R's Representation	$\{\hat{P}^{H}_{R}, P^{H}_{\hat{H}_R}, \hat{N}^{L}_{\hat{H}_R}, \hat{N}^{L}_{R}, N^{L}_{\hat{L}_R}, \hat{P}^{H}_{\hat{L}_R}\}$
	$\hat{P}^{H}_{R}$ – R's estimate of H's displayed positive affect.
	$P^{H}_{\hat{H}_R}$- R's estimate of H's conscious positive affect.
	$\hat{N}^{L}_{\hat{H}_R}$- R's estimate of H's responsiveness to L's negative emotions.
	$\hat{N}^{L}_{R}$- R's estimate of L's displayed negative affect.
	$N^{L}_{\hat{L}_R}$- R's estimate of L's conscious negative affect.
	$\hat{P}^{H}_{\hat{L}_R}$- R's estimate of L's responsiveness to H's positive affect

We consider three facets of H's mental state, $H = \{P_H^H, \widehat{P}_H^H, \widehat{N}_H^L\}$, that are indicative of H working to engage in a positive and open interaction. First, we must consider H's positive responsiveness toward L of which she is consciously aware (P_H^H). In a positive and open interaction, H wants to be attentive to L and show respect in situations where L shows weakness (be open to any negative emotions expressed by L) (e.g. Retzinger, 1991; Tickle-Degnen, 2006). Percepts such as sustained gaze toward the other, relaxed posture, and calm, organized speech can be indicative of this state.

Second, H also has an estimate of how she is conveying this positive responsiveness ($\widehat{P}_H^H$). H has a notion of the degree to which she is being attentive and uplifting toward L. The compassion H attempts to show L may vary from her true conscious state. H can try to appear open and attentive while not having any compassion for the other. Alternatively, H could be open and want to uplift the other but may not convey this behaviorally. Percepts that identify a disparity between verbal and nonverbal behavior (e.g. H averting her gaze while using uplifting/positive language) reveal a difference between conscious affect and displayed affect.

Finally, we consider H's understanding of L's negative thoughts about the self (attempts to disengage) ($\widehat{N}_H^L$). An understanding of the other is important if an individual is going to appropriately respond. Percepts such as facial mimicry or explicit naming of the other's emotion show H's desire to understand and be responsive toward the other.

We also define a partial theory of mind for L, $L = \{N_L^L, \widehat{N}_L^L, \widehat{P}_L^H\}$. This set consists of L's conscious negative thoughts about the self that drive a desire to disengage from the other (N_L^L). For example, if L is weak (e.g. experiencing unworthiness shame (Giner-Sorolla, 2012)) and being avoidant (alienating herself from the other) because she is struggling to deal with the situation, the relationship suffers (Bush & Folger, 2010a). Percepts such as being overly quiet or avoidant gaze can indicate this state.

Another piece of L's mental state is her estimate of how she is conveying these negative emotions ($\hat{N}^{L}_{L}$). One may conceal weakness or behave in an isolating way even if they do not wish to be alienated from the other. Again, percepts relating to a division between verbal and nonverbal behavior can help reveal this state. Finally, L's understanding of H's positive emotions also must be considered ($\hat{P}^{H}_{L}$). If L is going to be responsive to H, she must understand H's behavior.

To spot problems in the relationship between H and L, R must understand when the mental states and behaviors of H and L are problematic for long-term relationship health. R is also represented by a set, $R = \left\{\hat{P}^{H}_{R}, P^{H}_{\hat{H}_R}, \hat{N}^{L}_{\hat{H}_R}, \hat{N}^{L}_{R}, N^{L}_{\hat{L}_R}, \hat{P}^{H}_{\hat{L}_R}\right\}$. R maintains its own estimate of H's displayed positive affect ($\hat{P}^{H}_{R}$), an estimate of H's conscious positive affect ($P^{H}_{\hat{H}_R}$), and an estimate of H's understanding of L's negative emotions ($\hat{N}^{L}_{\hat{H}_R}$). It also has its estimate of L's displayed negative affect ($\hat{N}^{L}_{R}$), an approximation of L's conscious negative emotions ($N^{L}_{\hat{L}_R}$), and an estimation of L's understanding of H's positive affect ($\hat{P}^{H}_{\hat{L}_R}$).

Our representation of R allows us to define the problematic relationship states that must be recognized. The following subsection explicitly enumerates the relationship state space and presents a computational model through which a robotic agent can recognize the enumerated states where dissatisfaction exists (addressing subsidiary question 2). The second subsection enumerates behaviors that address each relationship state and an action-selection mechanism that supports a healthier relationship (subsidiary question 3).

A general overview of R's intervention algorithm is shown in Algorithm 3.1. R can have a variety of sensors (e.g. cameras, microphones, and physiological signal recorders) writing to circular buffers. It iteratively uses functions to abstract those sensor readings to usable percepts. The representations of the percepts are specific to the individual percepts; they can be scalars, categorical, real valued, etc. The percepts are iteratively mapped to a vector that indicates the problematic states that are active in the relationship at the current time. The data flow from the sensors to the time at which the active states is set is covered in section 3.1. The implementation of this part of the computational architecture is in 4.2.

The active problematic relationship states define what actions (from predefined sets) that the robot can take to support the relationship at a certain time. A coordination function is used to select a specific intervention before the robot enacts the intervention. The robot will continually choose interventions (which include doing nothing) to support the relationships based on what relationship states are currently active. The action selection is covered in 3.2 and 4.3. An exemplary situation of how the model could identify and intervene to support positive change is given in section 3.3. This example covers the data flow from the sensors to the robot intervening, including: the representation of percepts (abstractions of sensor data), the mapping of percepts to active states, as well as the selection of specific interventions for the robotic agent to enact.

Algorithm 3.1: General Robotic Intervention Algorithm

Intervention Algorithm

```
Define:
    Percept-Generating Functions F <f_1,..., f_m>
    Set States Function g   //Predefined mapping from percepts to states
    Behaviors B <β_empathy, … , β_maintain>
    Coordination Function C
Input:
    Pointers to Sensor Objects S <s*_1,...,s*_n>
    Pointers to Percepts P <p*_1,...,p*_m>
    Pointers to State Vector X*

//Step 1: Identify Active Relationship States
For each s_i ∈ <s_1,..., s_n>
    s_i->start( )
For each f_i ∈ <f_1,..., f_m>
    Spawn thread f_i(S, p_i)  //f_i(S)→ p_i iteratively
Spawn thread g(P, X)  //g(P)→ X iteratively

While TRUE //Intervene a certain number of times

    //Step 2: Choose Intervention  - note that the chosen intervention
    //           could be to do nothing

    Initialize M //Potential behavioral manifestations
    For each state x_i ∈ X
        If x_i == 1 //is active
            M.i = β_i(x_i) //a predefined set of behavioral
                           // manifestations to ameliorate state i
        Else
            M.i = ∅
        End If
    End For
    //Choose specific intervention (m_ik) – the intervention
    //selected depends on what the agent has already done
    m_ik ← C(M)

    //Step 3: Carryout Intervention
    Enact m_ik //The intervention m_ik could be to do nothing
    Wait(60) //Intervene and then wait 60 seconds
End While
Kill threads //End processes updating percepts and states
```

3.1 Recognizing States of Relationship Dissatisfaction

H and L's relationship may be in an acceptable state (where mutual satisfaction exists) or in one or more states of dissatisfaction. Let X be the relationship state space.

$$X = \{x_{insensitivity}, x_{negativity}, x_{intrapersonal\_discordance\_high\_positive}, x_{intrapersonal\_discordance\_low\_negative}, x_{interpersonal\_discordance\_high\_positive}, x_{interpersonal\_discordance\_low\_negative}, x_{acceptable}\}$$

This state space is directly tied to the representation of R given above. Each state is binary, either present or not. The mental states of the human actors exceed thresholds that are problematic or they do not. These states are grounded in literature from transformative mediation, occupation therapy, and conflict theory in the following chapter (see Chapter 4.1). Each relationship state is described in Table 3.2, which appears on the following page.

We introduce shorthand for each state in our state space, which is used from this point forward. Each state corresponds with the state in same position in X above.

$$X = \{x_{ins}, x_{neg}, x_{intra\_h,}, x_{intra\_l}, x_{inter\_h}, x_{inter\_l}, x_{acc}\}$$

R must be able to identify which states are active at a particular time t. Let S be a vector of n sensors that allow R to interface with the environment, and let P be a vector of m percepts, abstractions of sensor readings, that indicate the presence of certain states.

$$S = \begin{bmatrix} s_1 \\ \vdots \\ s_n \end{bmatrix} \qquad P = \begin{bmatrix} p_1 \\ \vdots \\ p_m \end{bmatrix}$$

A sensor $s_i \in S$ writes its reading (output intensity) (λ_i) at each time step (t_i) to a circular buffer ($s_i(t_i) \rightarrow \lambda_i$). A percept $p_i \in P$ is an abstraction of one or more sensor values; a certain percept can be binary, defined categorically, as a real value, etc.

Table 3.2: Description of the relationship states that the robotic agent can identify. These states are operationalized in Section 4.1. They are derived from the Transformative Mediation and Occupational Therapy literature as well as the relationship representation given at the beginning of Chapter 3.

Description of Problematic Relationship States

Relationship State	Description
$X_{\text{insensitivity}}$	H has little positive affect for the other, i.e. H is inattentive (e.g. averts her gaze) or demonstrates hostility (e.g. raises her voice) toward L. $P^{H}_{\hat{H}_R}$ is below a set threshold.
$X_{\text{negativity}}$	L is experiencing strong negative affect, i.e. is withdrawn from the other (e.g. uses few utterances) or pushes H away (e.g. glares at H). $N^{L}_{\hat{L}_R}$ is above a set threshold.
$X_{\text{intrapersonal\_discordance\_high\_positive}}$	H inappropriately expresses her positive affect for L. H may have profound empathy for L but may appear neutral or even hostile. H may use conflicting affective cues, curtail affective expression altogether, or express doubt in how she want to proceed. There is a difference above a threshold between $P^{H}_{\hat{H}_R}$ and $\hat{P}^{H}_{R}$.
$X_{\text{intrapersonal\_discordance\_low\_negative}}$	L inappropriately express her negative affect toward H. L may seem frustrated and/or combative toward H but be willing to engage with H. L may use conflicting affective cues, curtail affective expression altogether, or express doubt in how she want to proceed. There is a difference above a threshold between $\hat{N}^{L}_{R}$ and $N^{L}_{\hat{L}_R}$.
$X_{\text{interpersonal\_discordance\_high\_positive}}$	L is not receptive to H's positive affect toward her. L may not respond to positive affect with mimicry or smiling. L does not state that H is being compassionate. There is a difference above a threshold between $P^{H}_{\hat{H}_R}$ and $\hat{P}^{H}_{\hat{L}_R}$.
$X_{\text{interpersonal\_discordance\_low\_negative}}$	H does not respond to L's negativity. H does not show mimicry. H does not state that L is experiencing negative affect. There is a difference above a threshold between $\hat{N}^{L}_{\hat{H}_R}$ and $N^{L}_{\hat{L}_R}$.
$X_{\text{acceptable}}$	H is attentive and uplifts L, L is not overly negative, each member of the dyad is expressing herself accurately, and the dyad members are both responsive.

Let F = {$f_1,\ldots,f_m$} be a set of functions such that $f_i(S) \rightarrow p_i$ where $f_i \in F$ and $p_i \in P$; each function f_i maps a sensor's or sensors' readings that fall within a specified sliding window of time to a specific percept p_i. See Algorithm 3.2 for function f_i.

Algorithm 3.2: One of the m functions from the set F. It transforms the raw sensor data to percept i.

Function f_i

Input: Pointers to Sensor Objects S <$s_1^*,\ldots,s_n^*$>
Pointer to Percept i <p_1^*>

Initialize D //Holds raw data
While TRUE
    For each sensor s_j associated with percept p_i
        D.s_j = s_j->read_buffer()
    End For
    p_i->write(y_i(D)) //y_i - predefined mapping from D to p_i
End While

We also define a function g(P) → X, which maps the entire percept vector, which contains the current percept values at time t, to the state space, which states are active at time t. The algorithm for function g is shown in Algorithm 3.3. The data flow from the sensors to a vector indicative of which states are active are shown Figure 3.2. The algorithm and figure appear on the following page.

Algorithm 3.3: Function that continually reads percept vector and transforms that vector's values into binary values indicative of which of the seven relationship states are active.

Function g

Input: Pointer to Percept Vector P*
Pointer to State Vector X*

```
1. While TRUE
      //The relationship states are binary (1 is active, 0 is inactive)
      //X = {x_ins, x_neg, x_intra_h, x_intra_l, x_inter_h, x_inter_l, x_acc}
2.    Initialize activeStates = {1,1,1,1,1,1,1} //All states start active
3.    For each state xi ∈ X
4.          For each percept pi ∈ P related to state xi
                  //conditions_met() – lookup table – can state be
                  //active given percept value
5.                If conditions_met(xi, pi) == FALSE
6.                      activeStates.xi = 0
7.                End If
8.          End For
9.    End For
10.   X->write(activeStates)
11. End While
```

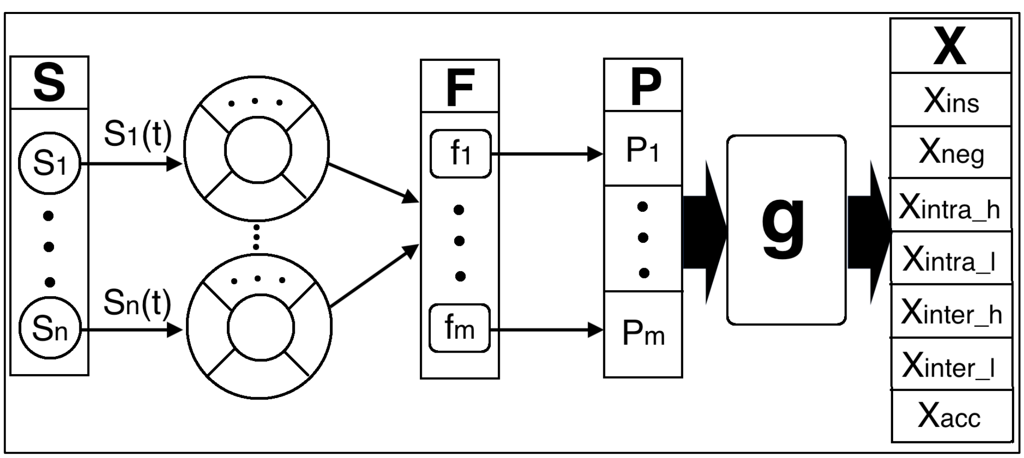


Figure 3.2: Data Flow from the sensors to the active states.

3.2 Action-Selection Mechanism to Support Relationship Satisfaction

Each state is addressed by a certain robot behavior. R must encourage empathy if H is insensitive. If L is overly negative, R must act to uplift L. When H or L's overt expression does not match her conscious feelings, R must encourage introspection so her conscious feelings match how they are conveyed. If a dyad member's understanding of the other is incongruent with the other's true state, R must reduce this incongruity. The association between relationship states and the behaviors are summarized in Table 3.3.

Table 3.3: Associations Between Relationship States and the Robotic Behaviors

Relationship State	Associated Robot Behavior
H is Insensitive to L ($\mathbf{x_{ins}}$)	Encourage Empathy in H
Pronounced Negativity in L ($\mathbf{x_{neg}}$)	Uplift L
Internal Discordance in H ($\mathbf{x_{intra\_h\_p}}$)	Encourage Introspection in H
Internal Discordance in L ($\mathbf{x_{intra\_l\_n}}$)	Encourage Introspection in L
Incongruence Between H's Feeling and L's Understanding of those Feelings ($\mathbf{x_{inter\_h\_p}}$)	Reduce the Incongruity Between H's Feelings and L's Understanding
Incongruence Between L's Feeling and H's Understanding of those Feelings ($\mathbf{x_{inter\_l\_n}}$)	Reduce the Incongruity Between L's Feelings and H's Understanding
Acceptable ($\mathbf{x_{acc}}$)	Maintain

B is a vector that contains R's available behaviors. Each of R's behaviors ($\beta_i \in \mathrm{B}$) can manifest in numerous ways. Let M be a vector which contains the sets of possible manifestations for each behavior such that $\beta_i(\mathrm{x_i}) \rightarrow \mathrm{M_i}$, where $\beta_i \in \mathrm{B}$, $\mathrm{x_i} \in \mathrm{X}$, $\mathrm{M_i} \in \mathrm{M}$, $\mathrm{M_i} = \{\mathrm{m_{i1}}, \dots, \mathrm{m_{ij}}\}$, and $\mathrm{m_{ik}} \in \mathrm{M_i}$ is specific manifestation of behavior i.

$$B(X) = \begin{bmatrix} \beta_{encourage\_empathy}(x_{ins}) \\ \beta_{uplift}(x_{neg}) \\ \beta_{encourage\_introspection\_h}(x_{intra\_h\_p}) \\ \beta_{encourage\_introspection\_l}(x_{intra\_l\_n}) \\ \beta_{reduce\_incongruence\_h\_pos}(x_{inter\_h\_p}) \\ \beta_{reduce\_incongruence\_l\_neg}(x_{inter\_l\_n}) \\ \beta_{maintain}(x_{acc}) \end{bmatrix} \qquad M = \begin{bmatrix} M_{encourage\_empathy} \\ M_{uplift} \\ M_{encourage\_introspection\_h} \\ M_{encourage\_introspection\_l} \\ M_{reduce\_incongruence\_h\_pos} \\ M_{reduce\_incongruence\_l\_neg} \\ M_{maintain} \end{bmatrix}$$

Currently, R chooses a single behavioral manifestation ($\mathbf{m_{ik}}$) from all of the possible ways of responding to current situation. A full example of how this is done is given in Section 3.3. Let C denote a coordination function such that $\mathbf{C(M)} \rightarrow \{\mathbf{m_{ik}}\}$. The coordination function maps the possible behavioral manifestations at the current time to a single behavioral manifestation (the response at that time). The coordination function is shown in Algorithm 3.4 (two pages below). We introduce shorthand for M that is used from this point forward.

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_{empathy} \\ \mathbf{M}_{uplift} \\ \mathbf{M}_{introspec\_h} \\ \mathbf{M}_{introspec\_l} \\ \mathbf{M}_{incongruence\_h} \\ \mathbf{M}_{incongruence\_l} \\ \mathbf{M}_{maintain} \end{bmatrix}$$

The coordination function begins by choosing one of the sets of M ($\mathbf{M_i} \in \mathbf{M}$). The set selected is the highest priority set as determined by some priority function. The priority depends on the importance of the behavior that is being enacted for the relationship. This could be set with the help of an expert. An exemplary priority function is given and applied in the full example given in Section 3.3.

The coordination function then chooses one behavioral manifestation ($\mathbf{m_{i,k}}$) from M_i probabilistically. The specific behavioral manifestation that is enacted by R is chosen using a weighted roulette wheel. The probabilities (weights) for the weighted roulette wheel are set by predefined probability functions. Each behavioral manifestation in the set M_i has a predefined probability function associated with it of the form shown in Algorithm 3.4 (on line 16). The probability of choosing a specific behavioral manifestation depends on the number of times that behavioral manifestation has been enacted as well as the number of times that the other behavioral manifestations in the set have been enacted. The probability functions will always sum to 1.

The probabilities of the weighted roulette wheel are updated each time the agent intervenes in the relationship. Any of the behavioral manifestations in the set with nonzero probability can be chosen. The definitions of the probability functions depend on the actual behavioral manifestations in a certain implementation. Exemplary probability functions are given in the full example in Section 3.3. The example also steps through how these probability functions are updated when the robot has intervened.

Algorithm 3.4: The coordination function.

```
                         Function C
Input: Potential Behavioral Manifestations M
Output: Behavioral Manifestation to Enact m_ik

//Step 1: Choose highest priority behavior set
1. maxPriority = 0  //Initialize maximum priority
2. maxSet  = M.1 //Max priority behavior set
3. For each set M.i ∈ M    //where M_i = {m_i1,...,m_iN}
4.     If priority(M.i) > maxPriority //a predefined lookup table
5.         maxPriority = priority(M.i)
6.         maxSet = M.i
7.     End If
8. End For

//Step 2: Choose specific behavioral manifestation
9.  Initialize weight = {0,...,0}
10. For each m_ij ∈ M.i
11.     weight.m_ij = probability(m_ij) //lookup table - probability
12. End For
    //Choose intervention using weighted roulette wheel
13. m_ik = weighted_roulette_wheel(weight)

//Step 3: Update probability lookup table values
14. count(m_ik) = count(m_ik) + 1 //times m_ik has been enacted
15. For each m_ij ∈ M.i
      //update values in lookup table with predefined
      //probability distribution, p(m_ij|count(m_i1),...,count(m_ij))
16.   probability(m_ij) = p(m_ij|count(m_i1),...,count(m_ij))
17. End For

//Step 4: Return Intervention
18. Return m_ik
```

After C chooses specific behavioral manifestation, the probability functions are updated. R carries out the chosen behavioral manifestation. If the behavior this intervention is meant to address remains active, and the behavior set associated with this behavior is still the highest priority, a specific behavioral manifestation is chosen from the set using the updated probabilities.

Finally, it makes sense to have a cutoff for the agent such that if a behavior remains active after a certain number of interventions, the agent returns to its unengaged behavior. The agent should not continue to disrupt the human-human relationship if its interventions do not seem to be helping the relationship. The next section gives a simple example of how this computational architecture works. The implementation of this architecture is given in Chapter 4 (see sections 4.2 and 4.3).

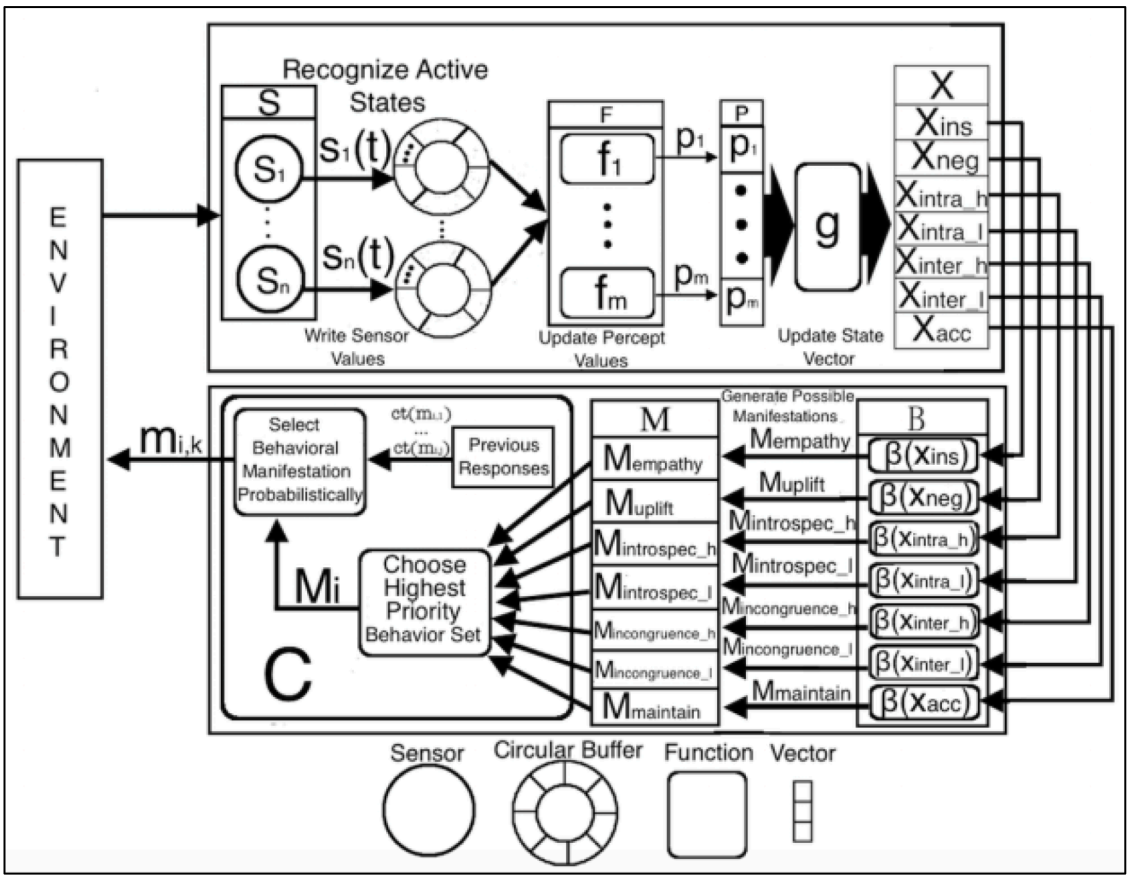


Figure 3.3: The data flow from sensors to intervention selection.

3.3 Example Implementation

A caregiver is speaking to the Parkinson's patient for whom she is caring. The Parkinson's patient is L (the lower-power dyad member). The caregiver is H (the higher-power dyad member).

Step 1: Sensors Writing to Circular Buffers

In this example, there are two sensors, a microphone on the caregiver's collar (s_{mic}) and an Empatica Wristband[1] on the caregiver's right wrist (s_{band}). The microphone (s_{mic}) writes a voltage measurement (λ_{mic}) in millivolts mV at each time step (t_{mic}) to a circular buffer. The voltage (λ_{mic}) depends on the input sound pressure, which depends on the loudness of the caregiver's voice.

$$s_{mic}(t_{mic}) \rightarrow \lambda_{mic}$$

The Empatica Wristband (s_{band}) is measuring the skin conductance level of the caregiver. The wristband outputs a conductance measurement (λ_{band}) in microsiemens (μS) at each time step (t_{band}) to a circular buffer. An individual's skin conductance level is associated with the emotional arousal of that individual and is often used as a measurement of stress (Bradley & Lang, 2000).

$$s_{band}(t_{band}) \rightarrow \lambda_{band}$$

In this example, both circular buffers hold the last five device readings. The input sound pressure level, which depends on the patient's voice, caused the microphone to output 20.77mV, 25.13mV, 22.22mV, 24.99mV, and 25.12mV into the circular buffer at the five most recent time steps. The skin conductance levels output by the band were:

---

[1] https://www.empatica.com/e4-wristband

.24µS, .22µS, .23µS, .20µS, and .23µS. The buffers are populated as shown in Figure 3.4.

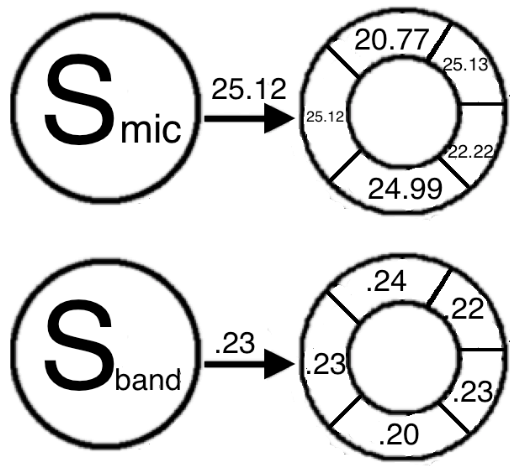


Figure 3.4: The two sensors affixed to the caregiver populate circular buffers of size 5.

Step 2: Sensor Readings Mapped onto Percepts

There are two percepts in this example, a percept indicative of whether or not the caregiver is yelling ($p_{yelling}$) and a percept indicative of the caregiver's stress level ($p_{stressed}$). These two percepts are defined by integer values. See Table 3.4.

$$p_{yelling} = \{p | p \in \{0, 1, 2, 3\}$$

$$p_{stressed} = \{p | p \in \{0, 1, 2\}$$

Table 3.4: The percepts, the range of potential numeric values, and the definitions corresponding to each value.

Definition of Possible Percept Values		
Percept	Value	Definition
$p_{yelling}$	0	The caregiver's voice is calm.
	1	The caregiver's voice is raised.
	2	The caregiver's voice is loud.
	3	The caregiver is screaming.
$p_{stressed}$	0	The caregiver's skin conductance is at baseline (indicating the caregiver is calm).
	1	The caregiver's skin conductance is raised (indicating some stress).
	2	The caregiver's skin conductance is high (indicating a lot of stress).

The value of each percept is iteratively set by a unique function. The function f_{yelling} maps the sensor values in the circular buffer from the microphone introduced above (s_{mic}) to the percept value p_{yelling}, and the function f_{stressed} maps the sensor values in the circular buffer from the wristband introduced above (s_{band}) to the percept value p_{stressed}. The values used in the mappings below were chosen ad hoc for the purpose of the example.

$$f_{\text{yelling}}(s_{\text{mic}}) = \begin{cases} 0, if \text{ median } \boldsymbol{\lambda}_{\textbf{mic}} \text{ over sliding window} \leq 2.5\text{mV} \\ 1, if \text{ median } \boldsymbol{\lambda}_{\textbf{mic}} \text{ over sliding window} > 2.5\text{mV and} \leq 10\text{mV} \\ 2, if \text{ median } \boldsymbol{\lambda}_{\textbf{mic}} \text{ over sliding window} > 10\text{mV and} \leq 53\text{mV} \\ 3, \text{if median } \boldsymbol{\lambda}_{\textbf{mic}} \text{ over sliding window} > 53\text{mV} \end{cases}$$

$$f_{\text{stressed}}(s_{\text{band}}) = \begin{cases} 0, if \text{ median } \lambda_{\text{band}} \text{ over sliding window} \leq .30\mu\text{S} \\ 1, if \text{ median } \lambda_{\text{band}} \text{ over sliding window} > .30\mu\text{S and} \leq .50\mu\text{S} \\ 2, if \text{ median } \lambda_{\text{band}} \text{ over sliding window} > .50\mu\text{S} \end{cases}$$

In the example, the caregiver is talking loudly. The median value held within the circular buffer is 22.22mV. Therefore, the function (f_{yelling}) sets the value of p_{yelling} to be 2. The median value for the caregiver's skin conductance (.23µS) falls in the calm range. Ranges for the stressed percept in an actual implementation would be determined using a baseline skin conductance level; this would be gathered before the interaction. The function f_{stressed} sets the p_{stressed} value to 0. The flow of the sensor data from the circular buffers to the current percept values is shown in Figure 3.5.

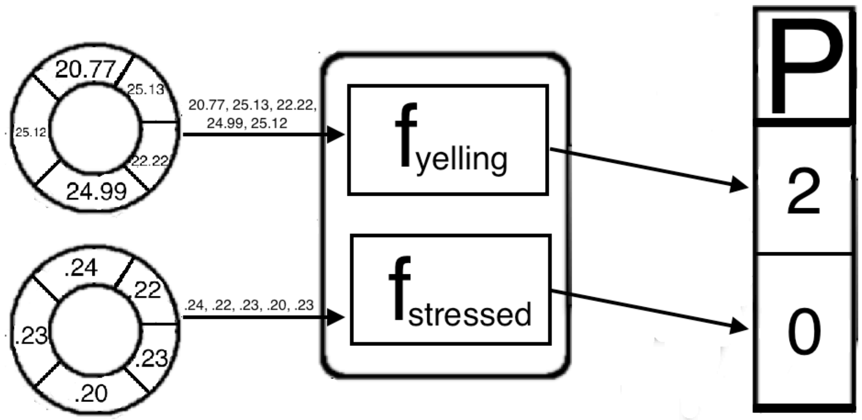


Figure 3.5: The data flow from the circular buffers containing the sensor data to the percept values.

Step 3: Mapping Current Percept Values to Active Relationship States

The two percepts defined in the previous subsection relate to the caregiver being insensitive (x_{ins}) (e.g. the caregiver could be screaming at the patient) and the caregiver inaccurately representing her feelings ($x_{intra\_h\_p}$) (e.g. the caregiver may be calm but speaking with a loud voice, implying she is upset). To simplify this example, we assume that these are the only two states relating to relationship dissatisfaction that can be active. That is, either one or both of these states ($x_{ins}, x_{intra\_h\_p}$) are active or just x_{acc} is active (if the caregiver is not insensitive or improperly presenting her internal state, then the relationship is assumed to in an acceptable state).

The function g maps the percept values contained in the vector P to a vector X that contains seven binary values that indicate which state(s) is/are active. If the caregiver's voice is not loud ($p_{yelling} < 2$), and he is calm ($p_{stressed} = 0$), then the acceptable relationship state is active. This state is active because the problematic relationship states are inactive.

If the caregiver is loud ($p_{yelling} \geq 2$) and at least mildly stressed ($p_{stressed} > 0$), then only insensitivity state is active. The caregiver is being insensitive by speaking to patient loudly and is expressing true frustration (i.e. the physiological measurement is consistent with the speaking volume). In this example, the caregiver's speaking volume is inversely related to positive affect for the patient. So larger values for this percept indicate lower positive affect for the patient. If the value of this percept greater than two, then the caregiver's positive affect for the patient falls below the acceptable threshold.

If the caregiver is not loud ($p_{yelling} < 2$) but is stressed ($p_{stressed} > 0$), then just the internal discordance state is active. The caregiver is concealing his stress and frustration. The magnitude of positive affect for the patient is high. The magnitude of the internal positive affect is low. There is a problematic difference between the magnitudes of these mental state components.

Finally, if the caregiver is loud ($p_{yelling} \geq 2$) but calm ($p_{stressed} = 0$), then both the insensitive state and the internal discordance state are active. The caregiver is acting insensitively and showing frustration (by speaking at a volume that exceeds the set magnitude) but is actually calm/relaxed. There is a difference between the magnitude of positive affect for the patient (which is low) and the internal positive affect, which is high.

$$g(P) = \begin{cases} \{0,0,0,0,0,0,1\}, & \text{if } p_{yelling} < 2 \wedge p_{stressed} = 0, \text{then } x_{acc} \text{ is active} \\ \{1,0,0,0,0,0,0\}, & \text{if } p_{yelling} \geq 2 \wedge p_{stressed} > 0, \text{then } x_{ins} \text{ is active} \\ \{0,0,1,0,0,0,0\}, & \text{if } p_{yelling} < 2 \wedge p_{stressed} > 0, \text{then } x_{intra\_h\_p} \text{ is active} \\ \{1,0,1,0,0,0,0\}, & \text{if } p_{yelling} \geq 2 \wedge p_{stressed} = 0, \text{then } x_{intra\_h\_p} \text{ and } x_{acc} \text{ are active} \end{cases}$$

Given that the caregiver is speaking loudly ($p_{yelling} = 2$) and is calm ($p_{stressed} = 0$), there are two states that are active ($X = \{1,0,1,0,0,0,0\}$). The caregiver is being insensitive, and there is discordance between the caregiver's feelings and how she is presenting herself. The data flow from the current percept values to the vector that represents which states are active is shown in Figure 3.6.

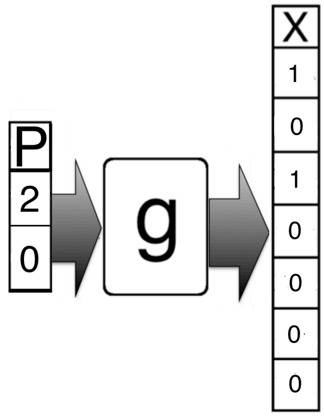


Figure 3.6: The data flow from the percept vector to the state vector.

Step 4: Generating Sets of Possible Actions for Each Active Behavior

The vector B contains the robot's potential behaviors. Each behavior is active if the state associated with the behavior is active. There are two active behaviors in the example, the encourage empathy behavior and the behavior to encourage introspection in the higher power individual in the relationship (the caregiver).

$$\mathrm{B(X)} = \begin{bmatrix} \beta_{encourage\_empathy}(x_{ins}) \\ \beta_{uplift}(x_{neg}) \\ \beta_{encourage\_introspection\_h}(x_{intra\_h\_p}) \\ \beta_{encourage\_introspection\_l}(x_{intra\_l\_n}) \\ \beta_{reduce\_incongruence\_h\_pos}(x_{inter\_h\_p}) \\ \beta_{reduce\_incongruence\_l\_neg}(x_{inter\_l\_n}) \\ \beta_{maintain}(x_{acc}) \end{bmatrix} = \begin{bmatrix} \beta_{encourage\_empathy}(1) \\ \beta_{uplift}(0) \\ \beta_{encourage\_introspection\_h}(1) \\ \beta_{encourage\_introspection\_l}(0) \\ \beta_{reduce\_incongruence\_h\_pos}(0) \\ \beta_{reduce\_incongruence\_l\_neg}(0) \\ \beta_{maintain}(0) \end{bmatrix}$$

Each behavior mapping generates a set of possible manifestations of that behavior. The encourage empathy behavior has three manifestations. The encourage introspection has one manifestation. The behavioral mappings $\beta_{encourage\_empathy}(x_{ins})$ and $\beta_{encourage\_introspection\_h}(\mathrm{x}_{intra\_h\_p})$ appear below.

$$\beta_{\mathrm{encourage\_empathy}}(\mathrm{x_{ins}}) = \mathrm{M_{empathy}} = \begin{cases} \{\mathrm{m_{empathy,1}}, \mathrm{m_{empathy,2}}, \mathrm{m_{empathy,3}}\} & if\ \mathrm{x_{ins}}\ active \\ \emptyset, & otherwise \end{cases}$$

$$\beta_{\mathrm{encourage\_introspection\_h}}(\mathrm{x_{intra\_h\_p}}) = \mathrm{M_{introspec\_h}} = \begin{cases} \{\mathrm{m_{introspec\_h,1}}\}, & if\ \mathrm{x_{intra\_h\_p}}\ active \\ \emptyset, & otherwise \end{cases}$$

The manifestations of these behaviors are explicitly defined below.

$m_{empathy,1}$ {<Nonverbal cues> [Nonverbal Cue Order]
[N1] Attend to the speaking party.}
$m_{empathy,2}$ {<Nonverbal cues> [Nonverbal Cue Order]
[N1] Attend to the speaking party.
[N2] Periodically glance at other party to "check in".}
$m_{empathy,3}$ {<Nonverbal cues> [Nonverbal Cue Order]
[N1] Attend to the speaking party.
[N2] Periodically glance at other party to "check in".
[N3] Reflect/Mimic affect of speaker.}

$m_{introspec\_h,1}$ {<Nonverbal cues> [Nonverbal Cue Order]
While high-power dyad member is speaking
[N1] Attend to high power dyad member.
[N2] Mimic affect of high-power dyad member.}

The vector M contains the sets of potential manifestations for all of R's behaviors. If a behavior is not active, then set of possible manifestations is empty. Therefore, M in this example is populated as shown below.

$$M = \begin{bmatrix} M_{empathy} \\ M_{uplift} \\ M_{introspec\_h} \\ M_{introspec\_l} \\ M_{incongruence\_h} \\ M_{incongruence\_l} \\ M_{maintain} \end{bmatrix} = \begin{bmatrix} \{m_{empathy,1}, m_{empathy,2}, m_{empathy,3}\} \\ \emptyset \\ \{m_{introspec\_h,1}\} \\ \emptyset \\ \emptyset \\ \emptyset \\ \emptyset \end{bmatrix}$$

Step 5: Choosing a Behavior to Enact

The coordination function begins by selecting the highest priority behavior set. In this example, we define a priority function such that nonempty sets of possible behavioral manifestations are given nonzero priority. Empty sets are given a priority of 0. The nonzero priority values were chosen ad hoc for this example.

$$\text{Priority}(\mathrm{M_i}) = \begin{cases} 0, \text{if } \mathrm{M_i} = \emptyset \\ 1, \text{if } \mathrm{M_i} \neq \emptyset \wedge (\mathrm{i} = \text{maintain} \vee \mathrm{i} = \text{introspec\_l} \vee \mathrm{i} = \text{incongruence\_h}) \\ 2, \text{if } \mathrm{M_i} \neq \emptyset \wedge (\mathrm{i} = \text{introspec\_h} \vee \mathrm{i} = \text{incongruence\_l}) \\ 3, \text{if } \mathrm{M_i} \neq \emptyset \wedge \mathrm{i} = \text{uplift} \\ 4, \text{if } \mathrm{M_i} \neq \emptyset \wedge \mathrm{i} = \text{empathy} \end{cases}$$

In the example, the only sets that are nonempty is $M_{emapthy}$ and $M_{introspec\_h}$. The $M_{empathy}$ behavior has a priority of 4. The $M_{introspec\_h}$ behavior has a priority of 2. Therefore, the set containing behavioral manifestations of the encourage empathy behavior is chosen.

$$\text{MAX}\begin{pmatrix} \text{priority}(\mathrm{M}_{empathy}), \text{priority}(\mathrm{M}_{uplift}), \\ \text{priority}(\mathrm{M}_{introspec\_h}), \text{priority}(\mathrm{M}_{introspec\_l}), \\ \text{priority}(\mathrm{M}_{incongruence\_h}), \text{priority}(\mathrm{M}_{incongruence\_l}), \\ \text{priority}(\mathrm{M}_{maintain}) \end{pmatrix} = \text{MAX}(4,0,2,0,0,0,0) \rightarrow \mathrm{M_i} = \mathrm{M}_{empathy}$$

Step 6: Choosing a Specific Behavioral Manifestation

The agent must now choose one of the three possible ways of enacting the encourage empathy behavior. This is done probabilistically; the behavioral manifestation is chosen using a weighted roulette wheel. Each behavioral manifestation has a probability function associated with it that depends on how the behavior has previously been enacted. These probability functions sum to 1. In the probability functions, $ct(m_{i,k})$ is the number of times the specific behavioral manifestation ($m_{i,k}$) has been enacted during the interaction.

$$\mathrm{p}\left(\mathrm{m_{empathy,1}} \middle| \mathrm{ct}(\mathrm{m_{empathy,1}}), \mathrm{ct}(\mathrm{m_{empathy,2}}), \mathrm{ct}(\mathrm{m_{empathy,3}})\right) = \frac{2 + \mathrm{ct}(\mathrm{m_{empathy,2}}) + \mathrm{ct}(\mathrm{m_{empathy,3}})}{5 + 2\mathrm{ct}(\mathrm{m_{empathy,1}}) + 2\mathrm{ct}(\mathrm{m_{empathy,2}}) + 2\mathrm{ct}(\mathrm{m_{empathy,3}})}$$

$$\mathrm{p}\left(\mathrm{m_{empathy,2}} \middle| \mathrm{ct}(\mathrm{m_{empathy,1}}), \mathrm{ct}(\mathrm{m_{empathy,2}}), \mathrm{ct}(\mathrm{m_{empathy,3}})\right) = \frac{2 + \mathrm{ct}(\mathrm{m_{empathy,1}}) + \mathrm{ct}(\mathrm{m_{empathy,3}})}{5 + 2\mathrm{ct}(\mathrm{m_{empathy,1}}) + 2\mathrm{ct}(\mathrm{m_{empathy,2}}) + 2\mathrm{ct}(\mathrm{m_{empathy,3}})}$$

$$p\left(m_{empathy,3}\middle|ct(m_{empathy,1}),ct(m_{empathy,2}),ct(m_{empathy,3})\right)$$

$$= \frac{1 + t(m_{empathy,1}) + ct(m_{empathy,2})}{5 + 2ct(m_{empathy,1}) + 2ct(m_{empathy,2}) + 2ct(m_{empathy,3})}$$

In the example, we assume that the agent has not enacted this behavior before. Therefore, we are computing the probability of choosing each behavioral manifestation given that the count of each manifestation is 0. When we plug the zeros into the probability equations given directly above, the agent has a forty percent chance of doing the first two interventions and a twenty percent chance of using the third intervention.

$$p(m_{empathy,1}|0,0,0) = \frac{2+0+0}{5+2*0+2*0+2*0} = .4$$

$$p(m_{empathy,2}|0,0,0) = \frac{2+0+0}{5+2*0+2*0+2*0} = .4$$

$$p(m_{empathy,3}|0,0,0) = \frac{1+0+0}{5+2*0+2*0+2*0} = .2$$

Let us assume the agent enacts the intervention labeled empathy 1. The agent attends to the party speaking for a certain number of exchanges. This attention might make the caregiver more self-aware and more aware of the patient; he may choose to change his interaction style.   Let us assume this intervention fails, i.e. the encourage empathy behavior remains active, and let us assume this behavior set is still the highest priority. The probability associated with each behavioral manifestation would be:

$$p(m_{empathy,1}|1,0,0) = \frac{2+0+0}{5+2*1+2*0+2*0} = \frac{2}{7} = .2857$$

$$p(m_{empathy,2}|1,0,0) = \frac{2+1+0}{5+2*1+2*0+2*0} = \frac{3}{7} = .4286$$

$$p(m_{empathy,3}|1,0,0) = \frac{1+1+0}{5+2*1+2*0+2*0} = \frac{2}{7} = .2857$$

Let us assume that the agent chooses the intervention labeled empathy 3. The agent would attend between the dyad members, check in with nonspeaking member to help frame interaction as a conversation between the two people, and mirror the affect of both dyad members to amplify the emotion in the room. After this intervention, the probabilities would be updated as shown below. The intervention labeled empathy 2 would have the highest chance of being chosen if the behavior is still active and the behavior set has the highest priority.

$$p\left(m_{empathy,1} \middle| 1,0,1\right) = \frac{2+0+1}{5+2*1+2*0+2*1} = \frac{3}{9} = .3333$$

$$p\left(m_{empathy,2} \middle| 1,0,1\right) = \frac{2+1+1}{5+2*1+2*0+2*1} = \frac{4}{9} = .4444$$

$$p\left(m_{empathy,3} \middle| 1,0,1\right) = \frac{1+1+0}{5+2*1+2*0+2*1} = \frac{2}{9} = .2222$$

Note that in between each intervention the robotic agent would be engaged in its unengaged behavior (defined in Chapter 1 and further discussed in Chapter 5) for a predefined period of time. This prevents from the agent intervening too much in the relationship. The agent also would only intervene a certain number of times. If the relationship members are not responding to the interventions, the agent should not continue to intervene and disrupt the relationship.

3.4 Conclusion

This chapter has introduced a general computational model and action-selection mechanism to help to support a positive and open interaction for long-term relationship health within a dyadic relationship with a power differential. We enumerated a partial theory of mind for a higher-power individual and a lower-power individual in a dyadic human relationship. The mental states that represented the human actors helped to reveal some of the problems that could arise in a dyadic human relationship and informed a representation of our peripheral robot.

The representation of the peripheral robot directly related to the six dissatisfying relationship states enumerated as part of our computational model. The first subsection, in addition to enumerating these six states, introduced a computational model to identify these states (addressing subsidiary question two from the first chapter). The relationship is represented as being in an acceptable state or in one or more states of dissatisfaction that need to be addressed by the robotic agent. The second subsection enumerated robotic behaviors that support the relationship when dissatisfying relationship states are present.

These behaviors are mapped to predefined response sets (the ways in which the behaviors can manifest). R chooses the highest priority behavior to enact according to a predefined priority function, which is defined using domain experts or literature. The specific manifestation of the behavior is picked probabilistically to allow for variation in the responses. This action-selection procedure addressed subsidiary question three.

The final section in this chapter provided a detail example of how the computational model could identify and intervene in a relationship between a Parkinson's patient and that patient's caregiver. The following chapter operationalizes the states identified here by providing explicit definitions from the literature, how practitioners identify them and implementation details for R's identification of these states. The following chapter also discusses R's behaviors to address these states, the specific implantation of behavioral manifestations and how these manifestations are chosen.

CHAPTER 4

COMPUTATIONAL ARCHITECTURE OPERATIONALIZATION AND IMPLEMENTATION

The question that drives this thesis, introduced in chapter one, asks how a robot can support positive change in hierarchical relationships that are entrenched in negative conflict. The computational model and action-selection mechanism introduced in the previous chapter work in conjunction to support more open and positive interactions in tiered relationships under such circumstances. This model identifies whether a relationship is in one or more states of dissatisfaction. The action-selection mechanism then chooses a predefined behavioral manifestation to address the relationship's highest priority issue. The positive change the robot supports in the relationship is helping the dyad move out of these six problematic states if and when they are present in the relationship.

In the second chapter, negative conflict was defined as a cycle of mutual weakness and alienation. Transformative mediators support empowerment and recognition shifts to help reverse this cycle and instead foster a constructive conflict cycle. As discussed below, the six states identified in the computational model are states of weakness and alienation. This means a positive change is an empowerment or recognition shift by the weak or alienated dyad member (e.g. Bush & Folger, 2010a). See Figure 4.1 for the division of the states.

State of Weakness	State of Alienation
• Negativity • Frustration/Withdrawal • Purposefully Not Engaging	• Insensitivity • Lack of Empathy/Hostility • Not Wanting to Understand
• Intrapersonal Discordance - L • Confusion/Uncertainty • Inability to Express Negative Affect	• Interpersonal Discordance - L • Not Understanding Positive Affect • Not Responding to Positive Affect
• Intrapersonal Discordance - H • Confusion/Uncertainty • Inability to Express Empathy	• Interpersonal Discordance - H • Not Understanding Negative Affect • Not Responding to Negative Affect

Figure 4.1: The six states of relationship dissatisfaction introduced as part of the computational model in chapter 3 are states of weakness and alienation defined in the Transformation Mediation literature. The states of weakness involve one individual not being able to successfully engage with the other. The states of alienation involve one individual not understanding or not even wanting to understand the other.

Transformative mediation operationalizes the attainment of empowerment and recognition by looking at shifts in the relationship members' language during the interaction (including body language) (Moen et al., 2001). Moen et al. (2001) explicitly identify types of dyad members' responses (with examples) that show "emerging empowerment" and "recognition". In this thesis, coding for emerging empowerment and recognition in the dyad members' language helps to identify positive change. See Appendix A for the rating scale used to identify emerging empowerment and recognition. Transformative mediation program directors have also used post-mediation questions to assess whether the practices helped to support empowerment and recognition (Folger, 2010). Post-interaction questionnaires are administered to participants to assess success.

The first subsection of this chapter walks through how these states are grounded in the literature and how they are identified in this thesis by the experimenters. Subsequently, in the second subsection, we provide the implementation details of the computational model used in the experiment discussed in chapter six. These subsections provide an answer regarding how the relationship is represented such that a robot knows when to intervene (subsidiary question two from chapter one).

The third subsection discusses how the robot supports positive change in the relationship. This includes an implementation of interventions (based upon the literature) for the experiment discussed in chapter six. The how the robot intervenes is related to answering subsidiary question three from chapter one.

Figure 4.2 shows the data flow of the entire computational architecture that was introduced in chapter three. Within this diagram, overlays have been added to show the subsections of the chapter in which that part of the architecture is discussed. The subsections follow the data flow. Subsection 4.2 introduces the sensors, how the sensor readings are abstracted into percepts (P) by a set of functions (F), and finally cover how these percepts are used to set the states (using a function g). Subsection 4.3 discusses the interventions that were implemented for the agent and how the robotic agent chose its specific interventions when the relationship was strained.

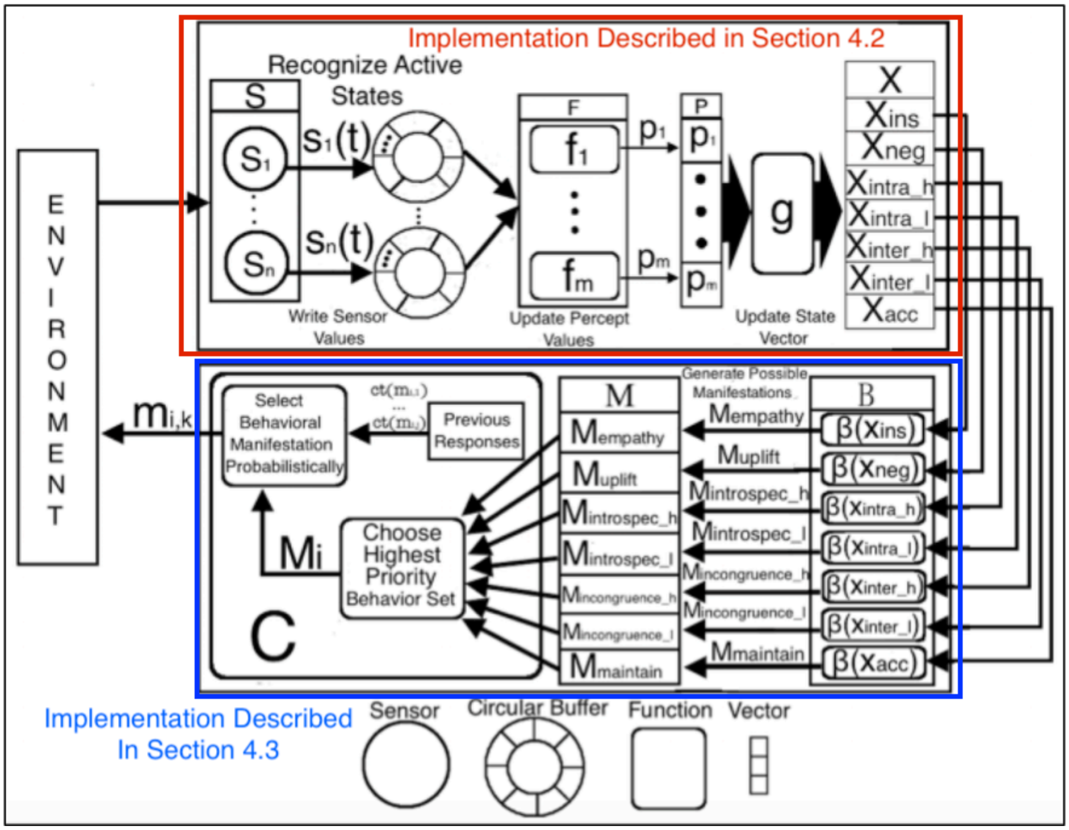


Figure 4.2: The data flow of the computational architecture described in chapter three. The second and third subsidiary sections provide the implementation details for this architecture. The second section gives details about identifying the problematic relationship states. The third section gives details about the robots interventions.

4.1 Operationalization of Computational Model

The proposed computational model identified six potentially co-occurring states of dissatisfaction in dyadic relationships. As discussed below, each state enumerated in the model is an instance when one dyad member shows signs of being weak or alienated from the other. Any such instance is an opportunity to support empowerment or recognition (Bush and Folger, 2010b; Moen et al., 2001).

Bush and Folger (2010b) as well as Moen et al. (2001) identify characteristic types of responses that show weakness. The first type identified responses that show negative affect such as frustration or anger (Moen et al., 2001). Bush and Folger (2010b) also cite displays of helplessness. When an individual begins to curse, makes proclamations about being unable to continue with the discussion, and/or explicitly references growing concern, the mediator knows the individual is weak and tries to support empowerment (Bush and Folger, 2010b; Moen et al., 2001). Such displays correspond with the state (Chapter 3) where the lower-powered individual is showing negative affect ($x_{negativity}$).

This state is operationalized looking for specific and explicit signs of disengagement from (or desire to end) the interaction. This disengagement can be in the form of withdrawal like symptoms (pulling away from the other person) or aggression/hostility towards the other (pushing away the other person). The individual can make explicit verbal statements that they do not want to continue with the discussion, cannot deal with the other person, or she may insult the other person. These verbal cues are often combined with nonverbal cues where the individual shows inattention toward the other, concealment from the other, or is domineering toward the other (Fig. 4.3).

The way the state was identified by human coders (a researcher and an individual without any knowledge of this research) appears in Appendix B under State of Weakness. The rating scales were drawn from transformative mediation (Bush & Folger, 2010b; Moen et al., 2001), conflict literature (Retzinger, 1991), and occupational therapy literature (Tickle-Degnen, 2006). This state of negativity is just one of three states of negativity that are identified in the computational model. It is validated as part of the experiment that is discussed in Chapter 6.

Figure 4.3: The negativity state sees the lower-power dyad member pulling away from or pushing away the higher-power dyad member through withdrawal or aggression. When humans decide to end direct engagement, whether through aggressive signaling or withdrawal, it is problematic. The left-hand side shows how aggression can push others away. Lions roar to signal other lions to stay away[2]. A person's raised voice may chase someone away similarly[3]. When people will not face each other and talk directly[4], they are withdrawn from one another and cannot effectively deal with interpersonal issues.

An artificial agent can begin to identify this state in an ongoing interaction using a microphone and a camera. In the following subsection, the implementation details of the computational model are discussed. The individual may speak very quietly (compared to a baseline) or have fragmented speech with frequent breaks and pauses showing inability to continue with the discussion as it is progressing. These speaking patterns are combined with gaze aversion; she will remove herself from the interaction by looking away from the individual who is trying to engage her. Alternatively, the individual could speak

---

[2] https://www.pexels.com/photo/nature-summer-yellow-animal-55814/
[3] https://www.pexels.com/photo/adult-anger-angry-angry-face-277870/
[4] https://www.pexels.com/photo/bench-nature-love-people-50592/

loudly (compared to a baseline), with a rapid tempo, and have a fixed gaze, feigning strength to cover the weakness (Moen et al., 2001).

Individuals are also in a state of weakness when they are experiencing intrapersonal conflict, i.e., there is conflict between individuals' internal states and the way in which they convey those states. Chapter 3 introduced two states of intrapersonal conflict ($x_{intra\_h\_p}$ and $x_{intra\_l\_n}$). The higher-power relationship member is unsuccessful at communicating her positive affect for/understanding of the lower-power relationship member ($x_{intra\_h\_p}$). The lower-power member is unsuccessful at expressing the negative affect she is experiencing ($x_{intra\_l\_n}$). When these states are active, individuals will often show a sense of confusion (Bush & Folger, 2010b; Moen et al., 2001). If a higher-power relationship member believes she is communicating empathically with a subordinate, and the subordinate does not respond as if this is the case, this can make the higher-power member question her competence and decision-making.

In Figure 4.4, two people are engaged in what appears to be a tense conversation. The dyad members are showing clear frustration toward one another. The man, however, has great compassion for the woman (as seen in the thought bubble). He wants to address concerns that she has, but he is not successful at expressing this to her. He is weak because he is not conveying himself successfully. His competency to successfully communicate and make decisions that will benefit the interaction is brought into question, and he does not know what to do.

There may be conflicting emotional cues expressed by the individual. For example, the man in Figure 4.4 can say out loud that he cares for the woman he is speaking to but his body language is confrontational. The individuals may not trust their decision-making or be unsure of how to proceed, and they may be explicit about this confusion (Bush & Folger, 2010b; Moen et al., 2001). Though they have desires internally to help and engage the other, they doubt that they are going to be able to do that successfully. This doubt and confusion about how to proceed makes the conflict stall.

Figure 4.4: This appears to be a tense conversation between two people[5]. Both dyad members show combative body language. The male, however, has high levels of empathy for the woman to whom he is speaking. He is in a weakened state because he is unsuccessful at conveying this empathy for her.

The identification of these states by an autonomous system requires an understanding of the individuals' conscious states. The agent would have to have advanced natural language processing capabilities to understand when individuals were expressing doubt or confusion about the situation/how to proceed. These states are not implemented as part of the autonomous system tested in this thesis for this reason.

---

[5] https://www.pexels.com/photo/man-and-woman-wearing-brown-leather-jackets-984950/

In negative conflict, weakness in one dyad member is driven by the other's self-absorption (Bush & Folger, 2010a). Table 4.1 summarizes the three states of weakness discussed above. The other three states that the computational model identifies in Chapter 3 are states of self-absorption or alienation. First, the transformative mediation literature describes situations where a dyad member is unable or unwilling "to give recognition" to the other dyad member (Bush & Folger, 2010b; Moen et al., 2001).

Table 4.1: Summary of the states of weakness, which are operationalized in the text.

State of Weakness	Definition	Cues Indicative of State	Implemented
Negativity	High Negative Affect – The lower-power individual acts to withdraw or push the other away. She limits engagement.	• Orientation Toward or Away from Partner (Camera) • Speech Tempo (Microphone) • Speech Volume (Microphone) • Speech Cohesion (Microphone)	Yes
Higher-Power Intrapersonal Conflict	Intrapersonal Discordance – The higher-power dyad member does not accurately convey her positive affect for the other member.	• Expression of confusion ("I don't know what to do") • Discordance between empathy/attention expressed nonverbally and verbally	No
Lower-Power Intrapersonal Discordance	Intrapersonal Discordance – The lower-power dyad member does not convey the negative affect she is experiencing accurately.	• Expression of doubt ("I don't think I can…") • Discordance between the individual's willingness to engage expressed nonverbally and verbal reports	No

Such displays correspond to the state of insensitivity ($x_{insensitivity}$) in the computational model. The higher-power dyad member lacks positive affect for the lower-power dyad member. The way in which the state was identified by human coders (a researcher and an individual without any knowledge of this research) appears in Appendix B under State of Alienation. Appendix B references Appendix C, which provides insights into identifying this state during discussion sections in which the higher-power dyad member is not speaking. It is difficult for a robot to identify this state in exactly the same way as transformative mediators, as this would require complex natural language processing. Transformative mediators look for individuals' paying "lip service" to the other's points, assuming the worst motives of the other party or minimalizing and trivializing the points made by the other party (Moen et al., 2001). The occupational therapy literature affords a means by which this state can be identified.

Literature in occupational therapy describes "suboptimal experiences" of rapport where there is a "failure to develop or maintain bond" between relationship members (Tickle-Degnen, 2006). The occupational therapy literature finds "suboptimal experiences" of rapport are clearly expressed in the nonverbal behavior of a disinterested or anxious dyad member (Tickle-Degnen, 2006). The dyad member is unfocused and hypo-responsive to his/her partner or shows vigilant attention and hyper-responsive to his/her partner (Tickle-Degnen, 2006).

The hypo-responsive partner is bored by the concerns of her partner and treats the partner as if she is irrelevant. Similar to paying lip service to a partner or minimizing the other's points (Moen et al., 2001), being unfocused shows a lack of any real consideration for what she is saying or how she feels. A robot could identify disinterest in the higher-power dyad member by noting that the higher-power dyad member often has an averted gaze when the lower-power dyad member is speaking. Additionally, the higher-power dyad member may not give verbal feedback while the lower-power dyad member is speaking.

The occupation therapy literature describes hyper-responsive partners as "anxious" with an "overriding concern about self" (Tickle-Degnen, 2006). These partners believe that their partners are not accepting of or not responsive to them, and they behave in a way that is "effortful and intense". This is much like the hostility or "demonization" described in the Transformative Mediation literature as an individual becomes increasingly alienated from the other (Bush & Folger, 2010a). The overwhelming concern about the self is growing weakness that leads the individual the separate herself from the other. The alienated person does not even want to understand the other. Prototypical examples of this state can be seen in Figure 4.5.

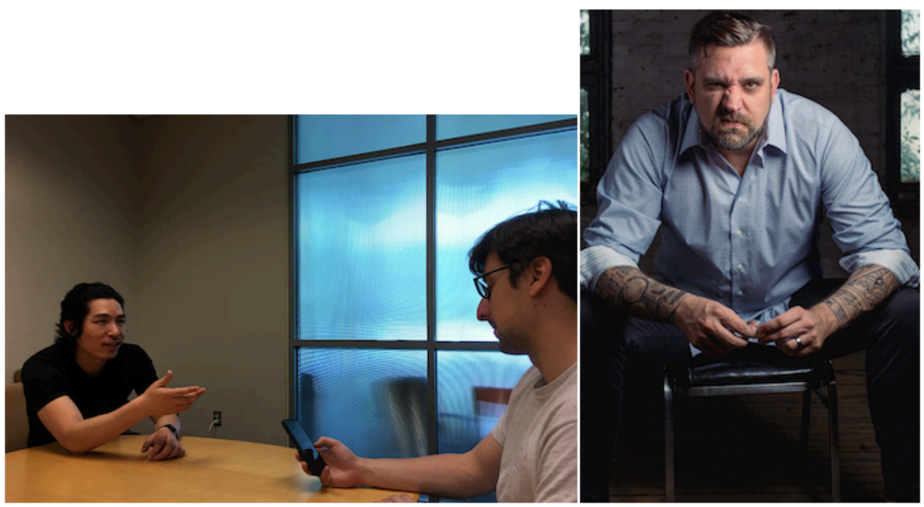


Figure 4.5: Higher-power individuals can aggressively reject the viewpoints and ideas of the lower-power individual or simply be bored and give the ideas no consideration[6].

[6] https://www.pexels.com/photo/man-sitting-on-black-leather-padded-chair-1447424/

A robot could identify a hostile dyad member by noting a fixed gaze and aggressive posture (e.g. leaning toward the partner). It could identify intense or harsh verbal and nonverbal behaviors that are destructive to social bonds/alienating (e.g. loud, rapid, and fragmented speech) (Moen et al., 2001; Retzinger, 1991; Tickle-Degnen, 2006). The robot can use a microphone and camera to look for the averted posture and lack of verbal feedback to identify a hypo-responsive partner or the aggressive posture and loud speech to identify a hyper-responsive partner. These are indicative of the presence of the insensitivity state.

Finally, there may be instances when one party is engaging with the other party, but she is not recognizing the viewpoints of the other party ($x_{inter\_h\_p}$, $x_{inter\_l\_n}$). Moen et al. (2001) describes how showing recognition to another individual's viewpoint does not mean to simply agree with that individual. It means working through the ideas of that other party. The party may acknowledge new pieces of information presented by her partner and may consider the circumstances under which she would agree with these viewpoints or admit to identifying with certain pieces of what is being said by the other party. The higher-power dyad member may listen to the lower-power dyad member but not fully appreciate to the frustrations the lower-power member is expressing. Similarly, the lower-power member may listen to the higher-power member without hearing her empathetic expressions.

When dyad members work through each other's ideas, it helps to reestablish a connection between the two parties. There is evidence that they understand each other or have some common nature / experiences (Bush & Folger, 2010a). It is problematic if the parties are just taking turns speaking to each other without engaging the each other's thoughts and feelings. The connection between the parties is never made; the parties remain disconnected from one another.

Identifying these states is difficult with an autonomous system that does not incorporate advanced natural language processing. The autonomous system would need to have an understanding of each party's internal state, and it would need to understand if the other party was addressing or being responsive to that internal state. For example, consider a situation when the lower-power dyad member was experiencing negative affect. The system would have to identify whether or not the higher-power dyad member was responsive to that negative affect. It would need to understand if the higher-power member was showing understanding toward these frustrations or was asking questions to understand where these frustrations were coming from, etc.

In a back and forth discussion, one party may or may not make explicit requests for recognition (Moen et al., 2001). In cases where the individual has made request for recognition, her partner has the option of showing recognition or explicitly refusing recognition. The refusal of recognition in this case would be an instance of the insensitivity state (x_{ins}) being active. The individual is explicitly affirming that she does not want to understand her partner by refusing to consider her thoughts and feelings when her partner is making explicit that she wants her to try and understand. The three states of alienation that have described here are summarized in Table 4.2 below (next page).

The six states that have been described in this section are instances of weakness and alienation that may occur in a hierarchical relationship. They are drawn from the transformative mediation literature; they are instances in which professional mediators would act to support relationships involved in negative conflict. The next section discusses an implementation of an autonomous system that tries to identify the presence of two of these states (the negativity, x_{neg}, and the insensitivity, x_{ins}, states) in a free-flowing discussion between two people. This section as well as the subsequent section helps to answer the second subsidiary question from Chapter 1.

Table 4.2: The summary of the states of alienation that are operationalized in the text.

State of Weakness	Definition	Cues Indicative of State	Implemented
Insensitivity	Low Positive Affect – The higher-power individual is hypo- or hyper-responsive to her partner. The inattentive or intense behavior serves to isolate the lower-power member.	• Orientation Toward or Away from Partner (Camera) • Speech Tempo (Microphone) • Speech Volume (Microphone) • Speech Cohesion (Microphone)	Yes
Higher-Power Interpersonal Discordance	Interpersonal Discordance – The higher-power relationship member is not responsive to the shame/frustration shown by the lower-power member.	• When the lower-power member is showing evidence of shame or frustration, the higher-power dyad member does not name the emotion or ask questions to understand it.	No
Lower-Power Interpersonal Discordance	Interpersonal Discordance – The lower-power relationship member is not responsive to the empathy shown by the higher-power relationship member	• When the higher-power member is showing evidence of empathy, the lower-power member does not name the emotion and continues to request recognition.	No

Transformative mediation has shown to be an effective way for people to positively transform relationships immersed in negative conflict. The agent should intervene in instances that transformative mediators would if it is to enact the same type of positive change. This section operationalized the states as defined in the transformative mediation literature. The next section provides further insight into how an agent can begin to identify these states.

4.2 Partial Implementation of the Computational Model

In this thesis, two states are implemented as part of an autonomous robotic system, one state of weakness and one state of alienation. The state of weakness, negativity (x_{neg}), as described in the previous subsection and chapter, is active when the lower-power dyad member has strong negative affect. Specifically, the state is active when the lower-power dyad member is creating distance between herself and the higher-power dyad member. This could be done through pushing the other away (expressing frustration/rage) or pulling away (withdrawing).

The autonomous robotic agent is going to identify this state using the same types of features as transformative mediators. As described in Appendix B, there are nonverbal and paralinguistic indicators of this state including the lower-power dyad member: yelling/speaking loudly, speaking rapidly, speaking in a fragmented way, glaring, and averting her gaze/orientation. The two sensors used to identify these characteristic state features are a camera and a microphone.

The state of alienation that was implemented as part of this thesis was the insensitivity state (x_{ins}). This state is active when the higher-power dyad member lacks positive affect (e.g. empathy) for the lower-power dyad member. As described in the previous section, this includes situations where the higher-power dyad member is hostile/frustrated with the lower-power individual or is withdrawn from the lower-power individual. It also includes situations where the higher-power individual is not responsive to the lower-power individual (does not show attention when the lower-power individual is speaking by having an averted gaze and/or not giving feedback while the lower-power individual is speaking). The indicators of this state largely mirror those described for the x_{neg} state. In the case of the x_{neg} state the lower-power member is driving the separation. In x_{ins}, the higher-power dyad member is driving the separation between the two dyad members. Identifying this state uses a camera and microphone as well.

The remainder of this subsection walks through the implementation of these two states in an autonomous robotic system following the computational model introduced in Chapter 3. There are two sensors (a microphone and camera); the readings of these sensors were abstracted into 4 different percepts (voice tempo, voice loudness, speech cohesion, and an individual's orientation toward/away from her partner). The way in which the sensors are used and how the percepts are derived are described in detail below. The four different percepts values were used to identify the presence of the two states in the dyadic human relationship.

Using the notation established in Chapter 3, there is a set of two sensors for the robot. The vector of sensors for the agent is:

$$S = \begin{bmatrix} s_{mic} \\ s_{cam} \end{bmatrix}$$

The first sensor used to identify the state of weakness or alienation in the individual was a lapel microphone affixed to that individual[7]. The second sensor was a webcam that was set up in a discrete location across the room from the individual that gave a clear view of

[7] https://www.amazon.com/gp/product/B01AG56HYQ/ref=ppx_yo_dt_b_asin_title_o05_s00?ie=UTF8&psc=1

the individual's face when she was looking toward her partner in the study[8]. These sensors continually write to circular buffers; the data from these circular buffers is continually abstracted into percepts by a set of functions F.

The set of functions F is F = $\{f_{fragmented}, f_{tempo}, f_{loudness}, f_{orienation}\}$. Each function maps the sensor data to a specific percept. The vector of percepts is:

$$P = \begin{bmatrix} p_{fragmented} \\ p_{tempo} \\ p_{loudness} \\ p_{orienation} \end{bmatrix}$$

The values that these percepts can take on are shown in Tables 4.3, 4.5, 4.7, and 4.8 below. The functions that map to these percepts to these values are given in Algorithms 4.1 – 4.4 below. The implementation that was used for the autonomous robot (all of the code) can be found on the Mobile Robot Lab's website[9]. These four percepts were used because they are some of the clearest indicators that experts look for when looking for weakness and alienation in individuals involved in destructive interactions (e.g. Retzinger, 1991).

The fragmented percept relates to an individual having jerky speech patterns with irregular or frequent pauses. The individual may speak in starts and stop or insert short filler phrases like "uh". This can be an indication that an individual's thoughts are disorganized and the individual is ashamed (Retzinger, 1991). The agent identifies this by noting when the individual has very short utterances or many utterances per minute.

---

[8] https://www.amazon.com/gp/product/B074SV76GK/ref=ppx_yo_dt_b_asin_title_o03_s00?ie=UTF8&psc=1
[9] https://www.cc.gatech.edu/ai/robot-lab/nri_thesis/code/All_Code.zip

An utterance is demarcated using openSMILE (Eyben et al., 2013). The software sets the start of an utterance when the volume exceeds a threshold for a certain number of readings. The speaking turn ends when the volume falls below the threshold for a certain number of readings. There are 100 readings per second related to loudness.

Table 4.3: Defines the potential values of the fragmented percept.

Values of the Fragmented Percept	
Integer Value	What the Integer Value Represents
1	The individual is speaking cohesively with lengthy utterances or few utterances per minute.
2	The individual is showing some fragmentation in thinking. She has short utterances or many utterances per minute.
3	The individual's speech is very fragmented. There are many utterances per minute and/or very short utterances.

As shown in Algorithm 4.1 and Table 4.3, the fragmented percept can take one of three integer values. As the value increases, the degree to which the thinking is fragmented increases. The thresholds for the utterances per minute and utterance length that dictate these three values were hand tuned using training data collected as part of the baseline study described in Chapters 5. Human coders labeled and provided justification for whether or not video segments had active states of weakness and alienation. The values seen in cases with active weakness and alienation and fragmented thinking informed the thresholds for the autonomous agent's implementation.

Algorithm 4.1: The algorithm to transform microphone data to the fragmented percept.

Function $f_{fragmented}$

```
Input: Pointers to Sensor Object <s*_mic>
       Pointer to Percept <p*_fragmented>
//Initialize data buffers and thresholds.
Initialize raw_buffer, utterance_length_buffer, utterance_time_buffer
Define T1, T2, T3, T4, T5

//Read and transform raw data
Spawn thread {
  While TRUE
        Initialize temp_buffer, temp_buffer_2 //Holds transformed data
        raw_buffer = s_mic->read_buffer()
        //Transform raw data to utterance lengths using openSMILE
        temp_buffer = transform_to_utterance_length_in_seconds(raw_buffer)
        //Record the times that the utterances concluded
        temp_buffer_2 = record_times_utterances_ended(raw_buffer)
        //Write to put utterance lengths and times in buffers
        utterance_length_buffer->write_safe(temp_buffer)
        utterance_time_buffer->write_safe(temp_buffer_2)
End While}

//Set Fragmented percept to one of the integer values defined in Table 4.3.
Spawn thread {
While TRUE
        Initialize temp_buffer, temp_buffer_2
        temp_buffer = utterance_length_buffer->read_safe() //Read utterance
        temp_buffer_2 =  utterance_time_buffer ->read_safe() //related data

        //Compute average utterance length and utterances/minute
        average_utterance_length = compute_average_utterance(temp_buffer)
        utterances_per_minute = compute_utterances_per_min(temp_buffer_2)
        If  average_utterance_length > T1 or
              utterances_per_minute < T2  and average_utterance_length > T3
                p_fragmented->write(1) //Speaking cohesively
        Else If  average_utterance_length > T4 and utterances_per_minute < T5
                p_fragmented->write(2)  //Speaking slightly fragmented
        Else
                p_fragmented->write(3)  //Speaking very fragmented
        End If
End While}
Wait_Until_End( ) //Wait interaction's end is signaled to kill threads
Kill Threads //Kill the data transformation and percept setting threads.
```

These thresholds (T_1 to T_5) were hand tuned because there is a paucity of training data to create regression models for this percept. The values output by openSMILE for utterance length (and openSMILE's decision to call a sound an utterance) depended strongly on the environment, the settings within openSMILE itself, and the hardware used as part of the study. Therefore, the training data gathered by the experimenters was the only thing that could be drawn upon to decide the decision boundaries for the fragmented percept.

The openSMILE library takes as input the raw microphone signal. It samples the raw signal at 100Hz and compares it to a threshold. This threshold demarcates the microphone is receiving input (as opposed just reading noise). When the raw input exceeds the threshold for a certain number of readings, openSMILE marks the beginning of that stretch as the beginning of an utterance. When the raw input falls below that threshold for a sufficient number of readings, openSMILE marks the end of the utterance. The algorithm outputs the length (in seconds) of that utterance. The algorithm also outputs the number of seconds since its start to end of that utterance (this acts as a time stamp for the utterance). Table 4.4 gives example outputs for openSMILE and how that output is used to set the fragmented percept.

In the implementation, complete utterances were often several seconds long and generally ended after a pause that would conclude a sentence. For example, the sentences, "I think that concealed carry on the Georgia Tech campus should be banned. Concealed carry introduces opportunities for lethal mistakes on campus." could be two utterances that last three to four seconds each if they are said as written. If the sentences instead are said, "I think. that. concealed carry. on the Georgia Tech campus. should be banned.

Concealed carry. introduces opportunities. for lethal mistakes. on campus.”, then they can be nine utterances that are approximately a second or a fraction of a second. Cohesive speech, like the first example, is indicative of an empowered individual (someone who is comfortable and sure about what she is saying). Speech showing fragmented thinking, like the second example, is indicative of someone who lacks empowerment (Retzinger, 1991). These are important indicators of whether or not the negativity and insensitivity states are active.

Table: 4.4: This table gives example output of the openSMILE program. It shows how the data are used to set the fragmented percept.

Example – Setting the Fragmented Percept									
Output of openSMILE	Utterance Times (seconds since program began)								
	34	40	42	43	55	64	80	83	90
	Utterance Lengths (in seconds)								
	3.3	.7	.85	.65	3.7	6.5	.5	.6	3.1
Average Utterance Length	(3.3+.7+.85+.65+3.7+6.5+.5+.6+3.1)/9 = 2.21 seconds								
Utterances Per Minute	(9/(90-34))*60 = 9.64 utterances per minute								
Setting the Fragmented Percept	2, slightly fragmented thinking, see Algorithm 4.1 Average Utterance Length is below T_1 (2.25) and above T_4 (0.8). Utterances/Minute is above T_2 (7) and below T_5 (25)								

The algorithm considers the last nine utterances made by the individual. If the individual’s average utterance was longer than 2.25 seconds, then fragmented percept was set to a 1. If the individual said fewer than seven utterances per minute, and the average utterance exceeded one second, then the fragmented percept was set to 1. This is in place because an individual may give a few short utterances (over the course of a minute) that are feedback responses. The first example given above in the text would result in the percept being set to 1 assuming the previous seven utterances were similar in length because the average time of the utterances was sufficiently high.

If the average utterance length fell below 2.25 seconds but stayed above 0.8 seconds, and the number of utterances per minute stayed below 25, then the percept is set to 2. The thinking is slightly fragmented. The second example above would result in the percept being set to 3 (the thinking is fragmented). There would have been nine utterances over the course of just about ten seconds. This would mean that there were 54 utterances per minute when considering the last nine utterances. Anytime the average utterance fell below 0.8 seconds, or the number of utterances per minute went above 25, the percept is set to 3 (indicating the thinking of the individual is highly fragmented). If the thinking of the individual is highly fragmented, this is a sign that the negativity and insensitivity are active. Fragmented speech is a sign of shame; the individual is not successfully engaging with the other person (Retzinger, 1991).

The tempo at which a person speaks can also be indicative of negative affect during an interaction (e.g. Retzinger, 1991). Individuals who are experiencing shame will often speak very rapidly. The tempo percept can take on three different integer values, as the values increase, the pace at which the individual is speaking increases. This percept is again just abstracting data from a microphone that the individual is wearing. The software package openSMILE processes the raw data from that microphone to determine the fundamental frequency of the individual's voice (Eyben et al., 2013).

The fundamental frequency is used as an estimate of the number of syllables per second being spoken by that individual. The algorithm to assign the tempo percept considers a weighted average of the number of syllables spoken per second over a sliding window of utterances. The algorithm considered the last 10 utterances. The weights are how long the utterance is, longer utterances give a better indication of how fast the individual is speaking.

The percept is assigned based on thresholds of what is a calm pace, what is a fast pace, and what is a very fast pace. The thresholds used to assign the percept were again hand tuned using the training data collected as part of the study described in Chapter 5. The percept values are given in Table 4.5. The algorithm to assign the percept is given in Algorithm 4.2. The thresholds in this case were 4.3 and 3.6 syllables per second. Example output of openSMILE and how it used to set the percept is given in Table 4.6.

Consider an individual saying, "I think that concealed carry should be banned on the Georgia Tech campus". This sentence has 17 syllables. If the individual says this sentence in three seconds, then she is saying 5.667 syllables per second. Assuming this is similar in speed to the other nine utterances, the tempo percept would be set to 3 because the individual is talking very fast. If the individual says this in six seconds, then she is speaking at 2.8333 syllables per second. The percept would be set as a 1 because she is speaking at a calmer and more relaxed pace.

Table 4.5: Defines potential values for the tempo percept.

Values of the Tempo Percept	
Integer Value	What the Integer Value Represents
1	The individual is speaking at a relaxed pace.
2	The individual is speaking at an accelerated pace.
3	The individual is speaking at a very fast pace.

Algorithm 4.2: Transforms the microphone data to the tempo percept.

```
                              Function f_tempo
Input: Pointers to Sensor Object <s*_mic>
       Pointer to Percept <p*_tempo>
//Initialize data buffers and thresholds
1. Initialize raw_buffer, utterance_length_buffer, syllables_per_second_buffer
2. Define T1, T2
//Read and transform raw data
3. Spawn thread {
4. While TRUE
5.          Initialize temp_buffer //Holds utterance length data (in sec)
6.          Initialize temp_buffer_2 //Syllables per second per utterance data
7.          raw_buffer = s_mic->read_buffer()

            //Raw data to buffer containing utterance lengths using openSMILE.
            //See text for detailed explanation. Thread safe write to shared buffer.
8.          temp_buffer = transform_to_utterance_lengths(raw_buffer)
9.          utterance_length_buffer ->write_safe(temp_buffer)
            //raw data to buffer containing the syllables per second in each
            //utterance using openSMILE. See text for detailed explanation.
10.         temp_buffer_2 = transform_to_syllables_per_second(raw_buffer)
11.         syllables_per_second_buffer ->write_safe(temp_buffer_2)
12. End While}
//Set Tempo percept to one of the integer values defined in Table 4.5.
13.  Spawn thread {
14.  While TRUE
15.         Initialize temp_buffer, temp_buffer_2 //Utterance lengths and rates
16.         temp_buffer = utterance_length_buffer->read_safe() // reads
17.         temp_buffer_2 =  syllables_per_second_buffer ->read_safe()
            //Computes a weighted average of syllables spoken per second
            //Weights are the lengths of the utterance.
18.         weighted_average_syllables_per_second =
                          compute syllables_per_sec(temp_buffer, temp_buffer_2)
19.         If   weighted_average_syllables_per_second  > T1
20.                p_tempo->write(3) //Speaking very rapidly
21.         Else If   weighted_average_syllables_per_second  > T2
22.                p_tempo->write(2)  //Speaking quickly
23.         Else
24.                p_tempo->write(1)  //Speaking at relaxed pace
25.         End If
26. End While}
27. Wait_Until_End( ) //Wait interaction's end is signaled to kill threads
28. Kill Threads //Kill the data transformation and the set percept threads.
```

Table: 4.6: This table gives example output of the openSMILE program. It shows how the data are used to set the tempo percept.

Example - Setting the Tempo Percept										
Output of openSMILE	Tempo (syllables/second)									
	4.3	4.2	2.5	2.8	2.4	6.25	1.9	1.5	2	2.8
	Utterance Lengths (in seconds)									
	3.3	.7	.85	.65	3.7	6.5	.5	.6	3	3.1
Weighted Average - Tempo	$\frac{(4.3 * 3.3 + 4.2 * .7 + 2.5 * .85 + 2.8 * .65 + 2.4 * 3.7 + 6.25 * 6.5 + 1.9 * .5 + 1.5 * .6 + 2 * 3 + 2.8 * 3.1)}{(3.3 + .7 + .85 + .65 + 3.7 + 6.5 + .5 + .6 + 3 + 3.1)} = 3.8$									
Setting the Tempo Percept	2, slightly accelerated pace, see Algorithm 4.2 The weighted average of the previous ten utterances was 3.8. This is above 3.6 and below 4.3.									

The loudness of an individual's voice can be indicative of a negative mental state (weakness or alienation). A person may speak loudly or stress certain words when enraged, or they could verbally withdraw (going long periods of time without speaking or speak very quietly) when ashamed (Retzinger, 1991). The loudness percept for an individual can take on one of seven different integer values. These values are defined in Table 4.7. The individual can be loud or screaming, can be stressing words (indicated by short stretches of speech exceeding a certain volume), can be silent for an extended period or speak quietly. There is also a percept value allotted that indicates the individual has just started speaking after an extended silence (7). Extended silence is an indication that the individual has been withdrawn. The percept value is set to be 7 for ten seconds after an extended silence to indicate there was verbal withdrawal recently (even though the individual may be speaking at a volume that is appropriate now).

The algorithm that sets the Loudness percept (see Algorithm 4.3) uses openSMILE to transform the raw microphone data to loudness values (there are 100 loudness values per second). It assigns a baseline average and standard deviation for the individual using her first ten seconds of speech.

The algorithm then computes the average loudness of the individual's speech over a sliding window of time. A threshold is used to exclude periods of silence when the individual is not speaking. The algorithm keeps track of the amount of time the individual has not been speaking to assess whether or not the she is verbally withdrawing from her partner.

The algorithm also computes the number of stressed words during the sliding window of time. Stressed words are utterances that are well above the baseline volume; they are placed into one of three different groups. Loud streaks are utterances in which the average volume was over one and a quarter standard deviations above the baseline. Very loud streaks are utterances in which the average volume was two standard deviations above the baseline. Shouted streaks are at least three standard deviations above the baseline. The algorithm computes the number of each type of stressed utterance.

This data is given to a selection function that assigns loudness percept value. The selection function is simply a series of if statements (described immediately below) that assign the value based on thresholds that were once again hand tuned given the training data collected in the study described in Chapter 5. The loudness function percept is summarized in Algorithm 4.3.

If the individual does not make an utterance for 30 seconds, the algorithm sets the percept value to 6 to indicate that the individual has verbally withdrawn. The individual has not spoken or provided any kind of verbal feedback to her partner. If the individual has made some kind of utterance over the last thirty seconds, then the algorithm considers the volume of what has been spoken.

If there are multiple shouted utterances, then the percept value is set to 8. The individual is shouting words and phrases, which is a clear sign of frustration (Retzinger, 1991). The percept value 8 is reserved for cases when the individual is shouting. In this case, there is a clear separation between the volume used by the individual and what would be appropriate in a relaxed conversation. In cases where there are multiple loud and very loud utterances, the percept value is set to 4. The individual is stressing certain words and phrases, evidence of being angered or enraged (Retzinger, 1991). The percept value of 5 is not included. This value was going to indicate an individual was speaking very loudly (as opposed to "just loudly"). During the model training process, this additional complexity was not found to help results and so a single loud value was used.

If the average volume of the recent speech is more than one and a quarter standard deviations above the baseline, then the algorithm sets the percept value to 3 because the individual is speaking with a raised voice. When the average volume is more than one standard deviation below the baseline, the algorithm set the value of the percept to 1 because the individual is speaking quietly. The value of two was going to indicate that the individual was speaking at a very low level (whispering). The thresholds that were used to exclude noise from being counted as speech made it so this value would never be set. Therefore, two was not included in the final implementation.

Table 4.7: The potential values for the loudness percept.

Values of the Loudness Percept	
Integer Value	What the Integer Value Represents
0	The individual is speaking at a normal volume without stressing or whispering words.
1	The individual is speaking quietly (the average volume over the sliding window of time is below the volume range established at the beginning of the discussion). The value 2 was reserved to identify when the individual was whispering (speaking very quietly). The data collected did not include individuals whispering. This additional precision with respect to how quiet the individual was did not improve our results.
3	The individual is speaking loudly (the average volume over the sliding window of time is above the volume range established at the beginning of the discussion).
4	The individual is stressing certain words (the speaking volume of the participant rose above the baseline volume range during several utterances in the sliding window of time). The value 5 was reserved to allow for additional precision with respect to how much louder than the baseline these words were. During testing, however, this additional precision did not improve our results.
6	The individual hasn't spoken for an extended period of time (or given any verbal feedback over a baseline volume). They have withdrawn verbally.
7	The individual just started speaking after an extended silence.
8	The individual screamed or made multiple utterances that were far above baseline / normal volume.

Algorithm 4.3: Transforms the raw microphone data to the loudness percept.

Function $f_{loudness}$

```
Input: Pointers to Sensor Object <s*_mic>
       Pointer to Percept <p*_loudness>
//Initialize data buffers and thresholds
1. Initialize raw_buffer, loudness_buffer,
2. Initialize baseline //Set as an average value and std across participants
3. baseline_set = FALSE
//Read and transform raw data.
4. Spawn thread {
5. While TRUE
6.      Initialize temp_buffer
7.      raw_buffer = s_mic->read_buffer()
        //raw data to buffer containing openSMILE loudness values
8.      temp_buffer = transform_to_loudness(raw_buffer)
9.      loudness_buffer ->write_safe(temp_buffer) //Thread safe write
10.     IF size(loudness_buffer) == 1000 and baseline_set == FALSE
11.         UPDATE baseline //set for individual after 10 sec. of speaking
12.         baseline_set = TRUE  //used to normalize percept for individual
13.     END IF
14. End While}

//Set Loudness percept to be one of the integer values defined in Table 4.7.
15.  Spawn thread {
16. While TRUE
17.     Initialize temp_buffer
18.     temp_buffer = loudness_buffer->read_safe() //Read
        //Length of time since person has spoken or given verbal feedback.
19.     silence_length = compute_curr_silence_length(temp_buffer)
        //Average volume of everything the individual has spoken
20.     avg_speaking_volume = compute_avg_speaking_volume(temp_buffer)
        //Number of stretches over/under baseline volume
21.     stressed_words = compute_stressed_words(temp_buffer)

        //Thresholds (based on the baseline) are used to set percept value
22.     the_percept_value = select_ value(silence_length,
                                          avg_speaking_volume,
                                          stressed_words,
                                          baseline)
23.     p_loudness->write(the_percept_value)
24. End While}
25.  Wait_Until_End( ) //Wait interaction's end is signaled to kill threads
26. Kill Threads //Kill the data transformation and the set percept threads.
```

The orientation percept relates to the individual's lean and gaze during the interaction. The orientation of an individual toward or away and the distance from her partner can be a good indication of a person's engagement with or separation from her partner (Retzinger, 1991). An individual who is enraged will often glare at the other person (be oriented toward the other person) a large percentage of the time. An individual who is enraged may also lean forward (in an intimidating manner) toward the other person. An individual who is oriented away from the other (e.g. averting her gaze) may be withdrawing from the other person.

The orientation percept can take on one of three integer values. The individual can appear aggressive or frustrated with the person with whom she is interacting (1). She can seem to be withdrawing from the person with whom she is interacting (2). Finally, she can be comfortable or natural with the person with whom she's interacting (0). These values are summarized in Table 4.8 below. To assign these values for an individual, the algorithm estimates the individual's lean (upright or not) and orientation (toward or away her partner).

The algorithm assigns the orientation percept for an individual twice every second. It begins by capturing a frame from the camera. It locates the individual's face within the image as well as the prominent features on the individual's face (e.g. nose mouth, chin, etc.) using the openCV and dlib libraries (Bradski & Kaehler, 2008; King, 2009). The algorithm makes an assessment about how the individual is leaning.

This assessment is made based on the height of the prominent facial features in the current image compared to their locations during a period at the beginning of the interaction. During the first minute of the interaction, when the individual is assumed to be getting settled into the interaction and is relaxed, a program tracks the location of the facial features in the image to get a range of the typical locations of these features in the image. If the individual becomes upset and leans forward or slouches considerably (showing a lack of engagement/indifference toward her partner), the location of these features will be in the lower part/below the typical range.

The algorithm then uses a generic three-dimensional model of a head (along with the locations of the facial features in the image) to solve for the orientation of the individual's head. This uses the openCV function solvePNP (Mallick, 2016). The generic model of the head is assumed to be level and facing forward. The solvePNP function outputs a transformation matrix for the camera that allows us to approximately project the points from the model of the generic head onto the image such that those points align with the corresponding points in the image. The algorithm projects a point that is assumed to be directly in front of the nose onto the image; this point falls directly inline with where the nose is pointing. The length of the vector between the nose and this point grows with the degree to which the head is turned away from directly ahead. If the individual is looking at her partner, the angle of this vector will be between the angle created by drawing a vector between the nose and the right side of the partner's head and the angle created by drawing a vector from the nose to the left side of the partner's head. The algorithm locates the partner's head in the image using basic template matching. The algorithm uses the derived vector's angle and length to estimate whether the individual is oriented toward or away from her the partner (whether her nose is pointed in the general direction of the partner or not).

The algorithm assesses the lean and orientation of the individual for each frame; the frames are captured every half second. It writes these values to a circular buffer. The proportion of frames the individual is leaning back and oriented away from her partner is computed over a sliding window of time. The percept is assigned by applying thresholds to these values. Again, these thresholds were set using training data collected in the study described in Chapter 5. The algorithm for assigning the orientation percept is summarized in Algorithm 4.4.

The algorithm considers the 90 most recent frames captured by the camera (the most recent 45 seconds of the interaction). If an individual is leaning toward and oriented toward her partner in over eighty percent of the frames (i.e. in at least 73 of the last 90 frames she is both leaning forward and has her nose pointed at her partner), then individual is considered to be oriented aggressively toward her partner (the value of the percept is set to 1). If the individual is oriented toward her partner over ninety-five percent of the time (i.e. in at least 86 of the last 90 frames her nose is pointed toward the partner), then she is considered to have an aggressive orientation toward her partner (the value of the percept is 1).

If the individual is oriented away from the other for over eighty percent of the analyzed frames (i.e. in at least 73 of the 90 frames her nose is not pointing toward her partner), then the percept is set to be a 2. Similarly, if the individual is leaned back for over ninety-five percent of the frames, and is oriented away from her partner for at least sixty-five percent of the frames, then the percept is set to 2. The individual is withdrawn/not actively engaging with her partner.

Table 4.8: Possible values for the orientation percept. If an individual is oriented away from another for large portions of time, that individual may be inattentive or shameful. If the individual is oriented toward the other for long periods of time, she may be aggressive.

Values of the Orientation Percept	
Integer Value	What the Integer Value Represents
0	The individual has a natural gaze and orientation toward the other person.
1	The individual has an aggression gaze and orientation toward the other person (leaning forward and/or staring at the other person for large portions of the sliding window of time).
2	The individual is withdrawn from the other person (gaze/head's orientation is averted from the other for large portions of the sliding window of time).

Given the definitions of the four percepts above (fragmented speech, speech tempo, speech loudness, and orientation), there is a function g that decides when the negativity (x_{neg}) and insensitivity (x_{ins}) states are active and inactive. This function is just a series of selection statements. The problematic relationship states our robot is identifying are not defined by a single behavior or the appearance of a single cue. Transformative mediators look for the appearance of multiple cues within an interaction to assess whether or not problems exist within a relationship (Bush, 2010; Noce, 2010).

The function g sets the insensitivity and negativity states active independently. It is possible for neither, one or both of the states to be active. The function g considers each defined percept. If there are at least two percept values that indicate the presence of negativity, then negativity is active. The percept values for the state of negativity depend on sensor reading gathered from the lower-power dyad member. If there are at least two values that indicate insensitivity is active, then insensitivity is set active.

Algorithm 4.4: Transforms raw camera data to orientation percept

Function $\mathbf{f_{orienation}}$

Input: Pointers to Sensor Object <s^{*}_{cam}>
Pointer to Percept <$p^{*}_{orientation}$>

```
//Initialize data buffers and thresholds
1. Initialize raw_buffer, orientation_buffer, lean_buffer
//Read and transform raw data
2. Spawn thread {
3. While TRUE
4.        Initialize temp_buffer, temp_buffer_2
5.        raw_buffer = s_cam->read_buffer()
          //raw data to buffer containing a frame of the video feed
6.        temp_buffer = transform_to_frame(raw_buffer)
          //Compute whether or not the individual under consideration is
          //oriented away from the other person in this frame. Uses dlib and
          //openCV to find the pose of the individual's head and the back of the
          //partner's head in the image. As described in the text, these can be
          //used to estimate whether or not the individual is angled toward or
          //away from her interaction partner.
7.        temp_value  = compute_oriented_away_or_toward(temp_buffer)
8.        temp_value_2 = compute_lean_toward_back(temp_buffer)
9.        orientation_buffer ->write_safe(temp_value) //Writes to
10.       lean_buffer->write_safe(temp_value_2)       //shared buffers
11. End While}
//Set Orientation percept to be one of the integer values in Table 4.8.
12.  Spawn thread {
13.  While TRUE
14.       Initialize temp_buffer
15.       temp_buffer = orientation_buffer->read_safe()
16.       temp_buffer_2 = lean_buffer->read_safe()
17.       proportion_oriented_away = compute_looking_away(temp_buffer)
18.       proportion_lean_back = compute_ leaning_back(temp_buffer)
19.       If is_too_high(proportion_oriented_away,  proportion_lean_back)
20.              p_fragmented->write(2) //The person is withdrawn
21.       Else If  is_too_low(proportion_oriented_away, proportion_lean_back)
22.              p_fragmented->write(1)  //The person is intense (staring)
23.       Else
24.              p_fragmented->write(0)  //There is an appropriate orientation
25.       End If
26. End While}
27. Wait_Until_End( ) //Wait interaction's end is signaled to kill threads
28. Kill Threads //Kill the data processing and percept writing threads.
```

Table 4.9: Description of function g that maps percepts to active relationship states.

State	This State is Active If	Description
x_{neg}	One of the following holds for the lower-power relationship member.	The negativity state is active when the lower-power relationship member has high negative affect.
	$p_{loudness} = 8$ **and** ($p_{tempo} > 1$ **or** $p_{fragmented} > 1$ **or** $p_{orientation} > 0$)	Yelling and there is at least one other piece of evidence showing negative affect and low engagement. Signs of frustration.
	$p_{fragmented} > 1$ **and** $p_{loudness} = 1$	The individual's thinking has signs of being fragmented and is quiet. Signs of shame.
	$p_{loudness} = 1$ **and** $p_{orientation} = 2$ **and** ($p_{tempo} = 3$ **or** $p_{fragmented} = 3$)	Speaking quietly with a withdrawn orientation and is speaking quickly or in a fragmented way. Signs of withdrawal/shame.
	($p_{loudness} = 3$ **or** $p_{loudness} = 4$) **and** $p_{orientation} = 1$	Speaking loudly or stressing words. Orientation is aggressive. Frustration.
	$p_{loudness} = 1$ **and** $p_{tempo} > 1$	Speaking quietly and quickly. Shame.
	$p_{tempo} > 1$ **and** $p_{orientation} = 2$	Speaking quickly with withdrawn lean and gaze. Signs of shame.
	$p_{tempo} > 2$ **and** ($p_{loudness} = 3$ **or** $p_{loudness} = 4$)	Speaking very quickly and loudly or stressing words. Signs of shame/frustrations.
	$p_{tempo} = 2$ **and** $p_{fragmented} > 1$ **and** ($p_{loudness} = 3$ **or** $p_{loudness} = 4$)	Speaking loudly, quickly, and fragmented. Feigning strength to conceal shame.
	($p_{loudness} = 3$ **or** $p_{loudness} = 4$) **and** $p_{orientation} = 2$	Feigning strength while showing sings of avoidance. Signs of shame.
	$p_{tempo} = 3$ **and** $p_{fragmented} > 1$	Speaking very quickly and fragmented. Signs of shame.
	$p_{loudness} = 7$ **and** ($p_{tempo} = 3$ **or** $p_{fragmented} = 3$ **or** $p_{orientation} = 2$) **and** ($p_{tempo} > 1$ **or** $p_{fragmented} > 1$)	Individual has been verbally withdrawn. There is at least one strong verbal indicator of shame in previous utterances (very rapid speaking or very fragmented speaking) or several less strong indicators of shame (e.g. orientation and fast speaking).
x_{ins}	One of the following or any of those given above is true for the higher-power member.	This insensitivity state is active when the higher-power relationship member lacks positive affect/is not attentive/engaged.
	($p_{orientation} = 2$ **or** $p_{orientation} = 1$) **and** $p_{loudness} = 6$	This is the case when the higher-power individual is inattentive or uncomfortably vigilant. She is not verbally responsive and is glaring or not looking at her partner.
x_{acc}	Neither of the previous two states is active.	This state is active when there are no problematic states active.

The percept values for insensitivity state depend on the sensor readings gathered from the higher-power dyad member.

The function g is summarized in Table 4.9. It provides all of the selection conditions for the states. The function takes as input the percept values which are derived as described above. It sets the states if they meet the criterion given in the table. Once the states are set, the robot uses the action-selection mechanism to select a specific behavioral manifestation as described in Chapter 3. The following subsection describes implemented interventions and how they are selected.

4.3 Intervention Strategies and Their Implementations

The previous subsection provided a partial implementation for the robot's computational model that identifies two problematic relationship states. Once the robot has identified that there is a problem in the relationship, it needs to intervene in the relationship. The types of interventions that are necessary to support the relationship relates to the third subsidiary question from Chapter 1. This subsection gives an overview of the interventions that were implemented as part of the autonomous system in this thesis. The code are available on the Mobile Robot Lab website[10], just as the implementation of the computational model. The interventions of the autonomous systems are rooted in literature from transformative mediation; they apply the principles of how relationship-focused mediators respond to these states.

The interventions serve to support the relationship members such that they are able to move out of the states of weakness and alienation to states of empowerment and

---

[10] https://www.cc.gatech.edu/ai/robot-lab/nri_thesis/code/All_Code.zip

recognition (e.g. Jorgensen et al., 2001). Practitioners of transformative mediation stress the importance of not urging relationship members to act in a certain way. Instead, they stress amplifying the behavior and emotion in the room so relationship members are able to accurately understand their own presentations as well as have an opportunity to hear what the other party is saying/feeling from a neutral source. In addition, they amplify the fact that the parties are having a discussion between one another. It is important for the parties to recognize that they are talking to one another (not making monologues) and that they are in control of how they handle the discussion. They also try to invite parties to make their own decisions and to speak openly. This stems from a belief that if the relationship members feel that they are in charge of how the process proceeds and are not subject to restrictions in how they express themselves, then they are already gaining empowerment.

In Chapter 3, the action-selection mechanism began to choose a specific behavioral manifestation by choosing the robot's highest priority behavior. The robot did this by referencing a predefined priority function that ranked the robot's behaviors one to six. The priority function ranked the behaviors based on what is the most important relationship issue to address according to experts in transformative mediation, conflict theory, and occupational therapy. The robot enacts the active behavior that was ranked as having the highest importance.

In the experiments conducted (described in Chapters 5 and 6), the uplift behavior and the encourage empathy behavior were never active at the same time. The studies conducted involved a single participant who was placed in either a lower- or higher-power role in the relationship. A study confederate filled the other role in the relationship.

The robot was not analyzing the study confederate's perceptual data. When the confederate was playing the lower-power role in the relationship, the negativity state would never be active because the lower-power relationship member's percept values were not being set. Similarly, when the confederate played the higher-power role in the relationship, the insensitivity state would never be active because the higher-power relationship member's percept values were not being set. A priority function did not need to be defined because only negativity or insensitivity could be active in a single trial.

The agent had to choose a single behavioral manifestation (response by the robot) from the predefined set of responses when the behavior was active during a trial. This section describes the sets of manifestations (responses by the robot) for each behavior and the selection of the specific manifestation from each set.

When the negativity state is active, the lower power dyad member is experiencing strong negative affect. The lower-power dyad member needs to be uplifted. The uplift behavior maps to six different predefined manifestations. These manifestations are defined in Table 4.10. The algorithm used to select these interventions is shown in Algorithm 4.5. The implementations of the interventions that this algorithm enacts are included in Algorithms 4.6-4.9.

The robot's first response to support the relationship is just nodding and smiling at each party while she speaks. The robot does this for one speaking turn for each person. This is to encourage each party to speak openly and to help empower each party member. This is in an attempt to uplift (empower) the lower-power dyad member who is showing weakness, but it also treats the dyad members the same, which is a critical tenant of transformative mediation (Bush & Folger, 2010a). It is not siding with the lower-power dyad member, which could shame the higher-power dyad member.

Table 4.10: The Implemented Interventions for the Uplift Behavior

$\beta_{uplift,1}$ {<Nonverbal cues> [Nonverbal Cue Order] [N1] Attend to the speaking party. [N2] Smile at speaking party. [N3] Periodically nod at party to indicate attention. [N4] Attend to the other party when speaker changes. [N5] Smile at new speaker. [N6] Periodically nod at party to indicate attention.}
$\beta_{uplift,2}$ {<Verbal cues> [Verbal Cue Order] [V1] "Ahem" <Nonverbal cues> [Nonverbal Cue Order] [N1] Attend to the speaking party. [N2] Smile at speaking party. [N3] Periodically nod at party to indicate attention. [N4] Attend to the other party when speaker changes. [N5] Smile at new speaker. [N6] Periodically nod at party to indicate attention.}
$\beta_{uplift,3}$ {<Nonverbal cues> [Nonverbal Cue Order] [N1] Wait until there is a break in the discussion. [N2] Smile <Verbal cues> [Verbal Cue Order] [V1] "I'm really enjoying hearing both of your insightful ideas"}
$\beta_{uplift,4}$ {<Nonverbal cues> [Nonverbal Cue Order] [N1] Wait until there is a break in the discussion. [N2] Display Sad/Angry Face <Verbal cues> [Verbal Cue Order] [V1] "I've sensed some uncertainty/frustration in the discussion. Is that correct?"}
$\beta_{uplift,5}$ {<Nonverbal cues> [Nonverbal Cue Order] [N1] Display Sad/Angry Face [N2] Embody Shameful/Frustrated Bodily Pose [N3] Display Statically For 10 Seconds.}
$\beta_{uplift,6}$ {<Verbal cues> [Verbal Cue Order] [V1] "Ahem" <Nonverbal cues> [Nonverbal Cue Order] [N1] Display Sad/Angry Face [N2] Embody Shameful/Frustrated Bodily Pose [N3] Display Statically For 10 Seconds.}

Algorithm 4.5: This algorithm selects a specific robot response when the negativity state is active.

Intervention Selection Algorithm for Uplift Behavior

```
//The robot has six manifestations of the uplift behavior. We initialize
//a bool corresponding to each to false because they have not been enacted.
enacted_smile_and_nod = false
enacted_ahem_smile_and_nod = false
enacted_verbal_uplift = false
enacted_verbal_amplify_weakness = false
enacted_nonverbal_amplify_weakness = false
enacted_ahem_nonverbal_amplify_weakness = false
//Check to see if the state is negativity state is active. If the negativity state is
//active, then trigger the intervention. If it is not active, wait and check again.
While TRUE
        Initialize x_neg  //Indicator of whether or not negativity is active
        x_neg = state_var->safe_read() //Function g sets value before read
        //If the state is active trigger the next intervention. These
        // interventions are shown in Algorithms 4.6-4.9.
        If x_neg == active_frustrated or  x_neg == active_withdrawn
            If enacted_smile_and_nod == false
                do_smile_and_nod()  //See algorithm 4.6
                enacted_smile_and_nod = true
            Else If enacted_ahem_smile_and_nod == false
                do_ahem_smile_and_nod() //See algorithm 4.6
                enacted_ahem_smile_and_nod = true
            Else If enacted_verbal_uplift == false
                do_verbal_uplift() //See algorithm 4.7
                enacted_verbal_uplift = true
            Else If enacted_verbal_amplify_weakness == false
                do_verbal_amplify_weakness (x_neg) //See algorithm 4.8
                enacted_verbal_amplify_weakness = true
            Else If enacted_nonverbal_amplify_weakness == false
                do_nonverbal_amplify_weakness(x_neg) //See algorithm 4.9
                enacted_nonverbal_amplify_weakness = true
            Else If enacted_ahem_nonverbal_amplify_weakness == false
                do_ahem_nonverbal_amplify_weakness(x_neg) //See alg. 4.9
                enacted_ahem_nonverbal_amplify_weakness = true
            End If
            Sleep_for_seconds(60)
        Else
            Sleep_for_seconds(5)
        End If
End While
```

The second intervention is the same as the first but preceded by a short verbal utterance (essentially the robot says "ahem"). This short utterance is drawn attention to the robot. A pilot study conducted to test the autonomous robot, which is discussed in detail in Chapter 6, found that individuals often were unaware the robot intervened in any way when the robot made strictly nonverbal interventions. See Algorithm 4.6.

Algorithm 4.6: Smile and nod intervention. The first intervention when the negativity state is active. The second intervention is the same aside from the robot making an "ahem" sound to draw attention to itself.

Algorithm − do_smile_and_nod()

```
//Attend to the party that is speaking. Smile. Nod when that party pauses to
//encourage the individual to keep speaking. Repeat with the second party.
If lower_power_member_is_most_recent_speaker
        turn_head_to_lower() //Turn toward the lower-power member
Else
        turn_head_to_higher() //Turn toward the higher-power member
End If
smile()
While most_recent_speaker_is_same //The robot faces the party that most
                                  //recently spoke until the other party speaks
    Initialize speaking_or_paused //variable indicating whether
                                  //speaker is actively speaking
    speaking_or_paused = check_if_speaking()
    If speaking_or_paused == PAUSED
        Nod()
    End If
End While
turn_head_180() //Attend to the other dyad member
While most_recent_speaker_is_same //The robot remains facing the party
                    // that most recently spoke until the other party speaks
        Initialize speaking_or_paused //variable indicating whether
                                      //speaker is actively speaking
        speaking_or_paused = check_if_speaking()
        If speaking_or_paused == PAUSED
            Nod()
        End If
End While
reset_robot_to_default_state()
```

The robots third intervention has the robot telling the dyad that it is really enjoying hearing all of their insightful ideas ("I'm really enjoying hearing both of your insightful ideas"). This again aims at empowering the lower-power dyad member. It says that both members have been contributing and have good ideas to offer. This does not exclude the higher-power dyad member, which could potentially be shaming and further alienate the two. The implementation of this intervention is shown in Algorithm 4.7.

Algorithm 4.7: This algorithm implements the verbally uplift intervention. This is the third intervention when the negativity state is active.

Algorithm − do_verbal_uplift()

```
//The agent waits until neither member of the conversation is speaking.
//It says it is enjoying hearing both of their ideas to encourage both members
// to continue to speak openly.
1. Wait_until_pause_in_conversation()
2. Speak("I'm really enjoying hearing both of your insightful ideas."
3. smile()
4. wait_seconds(5)
5. reset_robot_to_default_state()
```

After the third intervention, the robot tries to amplify the emotion in the room to help the dyad members acknowledge and confront this emotion directly and together. If the lower-dyad member expresses the weakness he/she is experiencing clearly and directly, it can be an empowering experience.

The robot amplifies the emotion by first saying that it senses uncertainty or frustration during the discussion ("I've sensed some uncertainty/frustration in the discussion. Is that correct?"). It says frustration if the lower-power relationship member has been glaring at her partner and uncertainty otherwise. It asks the dyad if its assessment is correct. This is done to show the assertion is tentative and allow for the dyad members to choose whether or not they want to acknowledge the emotion. Transformative mediators are always tentative and allow for themselves to be corrected when intervening (Bush & Pope, 2002). The agent does not recognize the response of the participants. It asks this question to allow for the dyad members to discuss the emotion amongst themselves. The robot is there to point out the emotion and allow for the dyad to decide how to proceed in the discussion. The robot makes a sad or angry face immediately after the verbal intervention. It makes an angry face if it has said "frustration" in the verbal intervention, and it makes a sad face if it has said "uncertainty" in the verbal intervention.

The affective facial display was added to the intervention after an important finding in the pilot study discussed in Chapter 6. Individuals reported being unsure of the robots nonverbal behavior and felt singled out by the robots nonverbal behavior. Therefore, it was important to use simple verbal statements to help clarify nonverbal interventions. The verbal statement is about the emotion in the discussion in general. The nonverbal behavior that appears in parallel with this verbal statement is used in subsequent interventions to amplify the emotion in the room (and hopefully does not single out either relationship member because it was associated with the emotion of the room). The implementation of the fourth intervention is shown in Algorithm 4.8.

The final two interventions are nonverbal. In the fifth intervention, the robot displays a sad or angry face and embodies the corresponding emotion. If the lower-power dyad member has been showing signs of shame, the robot will raise its arm and cover its face. If the lower-power dyad member has been showing frustration, the robot will appear to puff out its chest. Pictures of the robot's nonverbal behavior during these interventions are shown in Figure 4.5. The sixth intervention copies the fifth with the exception of using the "ahem" to draw attention to the agent. This, again, amplifies the emotion to allow for the parties to acknowledge the emotion and work through it together. The implementation for these interventions is shown in Algorithm 4.9.

Algorithm 4.8: This algorithm implements the verbally amplify weakness intervention. This is the fourth intervention when the negativity state is active.

Algorithm − do_verbal_amplify_weakness (x_{neg})

```
//The agent waits until neither member of the conversation is speaking and says
//that it senses that there is uncertainty or frustration in the discussion. It says
//frustration if the lower-power member has been glaring and uncertainty
//otherwise. This is to bring the emotion in the conversation to the surface without
//judging it and allow for the dyad members to address it together.
1. Wait_until_pause_in_conversation()
2. If x_neg == active_withdrawn
3.          Speak("I've sensed some uncertainty in the discussion. Is that correct?")
4.          display_sad_face()
5. Else
6.          Speak("I've sensed some frustration in the discussion. Is that correct?")
7.          display_angry_face()
8. End
9. wait_seconds(5)
10. reset_robot_to_default_state()
```

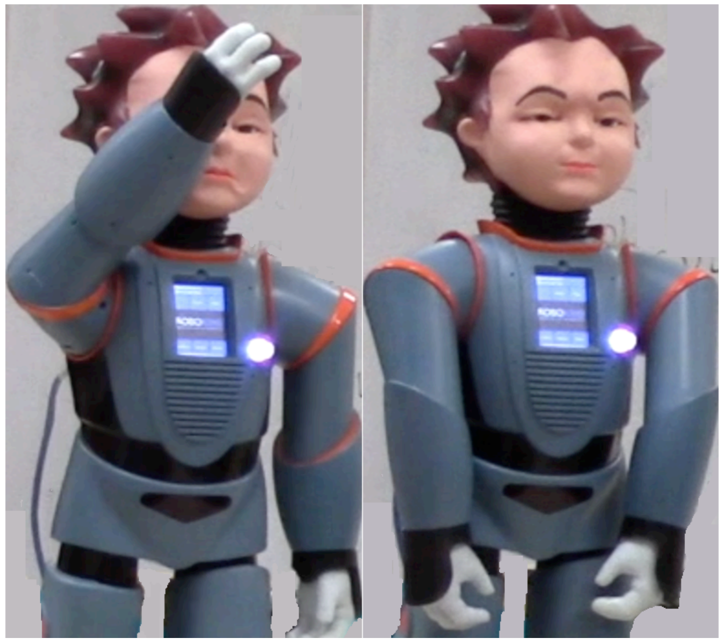


Figure 4.6: The shameful and frustrated displays for the uplift behavior's fifth and sixth interventions.

Algorithm 4.9: This algorithm implements the nonverbally amplify weakness intervention. This is the fifth intervention when the negativity state is active. The sixth intervention is the same with the robot first making an "ahem" sound to draw attention to itself.

Algorithm − do_nonverbal_amplify_weakness ($\mathbf{x_{neg}}$)

```
//The agent shows a prototypical expression of shame or frustration (face.   //and
body). This is to bring the emotion in the conversation to the surface //without
judging it and allow for the dyad members to address it together.
If x_neg == active_withdrawn
        display_sad_face()
        raise_arm_to_cover_face() //See Figure 4.6
Else
        display_angry_face()
        flex_arms_lean_forward() //Angry bodily posture, see figure 4.6
End
wait_seconds(5)
reset_robot_to_default_state()
```

As shown in Algorithm 4.5, the interventions employed by the autonomous robotic agent appear in the order in which they are described below, and each intervention is only triggered once (at most). Randomness was not incorporated into the choice of the behavioral manifestations because of three important things discovered when training the computational model (described in Chapter 5) and conducting the pilot study (described in Chapter 6). First, people often did not notice strictly nonverbal interventions. Interventions were kept as subtle as possible, so strictly nonverbal interventions were used first. If the behavior remained active, however, a simple utterance was used to garner some attention from the dyad members. This meant the "ahem" interventions followed the strictly nonverbal interventions.

As described in chapters five, triggering the interventions at the exact same time as a salient emotional expression was difficult. The nonverbal percepts that the robot used to identify the negativity and insensitivity states often did not occur in parallel with verbal statements that showed a desire to disengage from ones partner.

Therefore, it was important to reserve emotional amplification for instances when weakness and/or alienation had been witnessed multiple times in the interaction. This way the robot was pointing out something that was clearly felt by both dyad members. This meant that interventions were amplifying the emotions in the interaction should be after those that were about orienting the parties to each other or encouraging the dyad members to speak honestly.

Finally, in a five-person pilot study, which is further discussed in chapter six, certain participants were unsure about the nonverbal displays of the robot and certain participants were not comfortable with the robot addressing them directly (even when the robot was accurately perceiving negative emotion). On self-report measures, participants that had the robot address them directly rated the robot as not understanding what was going on in the interaction. This occurred even when it was intervening in the relationship at appropriate times (and trying to ameliorate real negative emotion). One participant went as far as to say, the robot "seemed to have no idea what we were saying to each other". When the robot addressed the confederate about him feeling emotional, and the robot was accurate, the participants rated the robot highly on its understanding of the situation.

Certain participants thought the robot was displaying an emotion felt by the speaker. When asked on self-report measures about the robot's interventions, there was a participant who said that when the robot made a sad face, he/she assumed that the robot thought that he/she was sad (but he/she did not feel sad). A verbal intervention was inserted into the fourth intervention to provide some insight into the robot's nonverbal emotional displays.

Given additional interventions more randomness could have been incorporated into the intervention ordering. The first implementation of this system was restricted to these interventions to prevent additional confounding variables from being introduced into the experiment described in chapter six. Limits on the number of interventions also allowed the experimenters to see how different participants responded to the same intervention.

The behaviors for the encourage empathy behavior are meant to support moves of recognition for the high-power dyad member. Again, for the same reasons explained above, these interventions were triggered in the order described below. These interventions are defined in Table 4.11 below. Algorithm 4.10 details the selection of the interventions of the robotic agent when the insensitivity state is active. Algorithms 4.11-4.14 show the implementation of the interventions.

Table 4.11: The Implemented Interventions for the Encourage Empathy Behavior.

$\beta_{empathy,1}$ {<Nonverbal cues> [Nonverbal Cue Order] [N1] Attend to the speaking party. [N2] Turn toward the listening party and back to the speaker every seven seconds. [N3] Attend to the other party when speaker changes. [N4] Turn toward the listening party and back to the speaker every seven seconds.}
$\beta_{empathy,2}$ {<Verbal cues> [Verbal Cue Order] [V1] "Ahem" <Nonverbal cues> [Nonverbal Cue Order] [N1] Attend to the speaking party. [N2] Turn toward the listening party and back to the speaker every seven seconds. [N3] Attend to the other party when speaker changes. [N4] Turn toward the listening party and back to the speaker every seven seconds.}
$\beta_{empathy,3}$ {<Nonverbal cues> [Nonverbal Cue Order] [N1] Wait until there is a break in the discussion. [N2] Smile <Verbal cues> [Verbal Cue Order] [V1] "I'm enjoying the chat you two are having"}
$\beta_{empathy,4}$ {<Nonverbal cues> [Nonverbal Cue Order] [N1] Wait until there is a break in the discussion. [N2] Display Sad Face <Verbal cues> [Verbal Cue Order] [V1] "I've sensed some discomfort during the discussion. Is that correct?"}
$\beta_{empathy,5}$ {<Nonverbal cues> [Nonverbal Cue Order] [N1] Attend to the speaking party. [N2] Turn toward the listening party and back to the speaker every seven seconds. When turning toward the listening party, the robot makes a sad face. [N3] Attend to the other party when speaker changes. [N4] Turn toward the listening party and back to the speaker every seven seconds. When turning toward the listening party, the robot makes a sad face.}
$\beta_{empathy,6}$ {<Verbal cues> [Verbal Cue Order] [V1] "Ahem" <Nonverbal cues> [Nonverbal Cue Order] [N1] Attend to the speaking party. [N2] Turn toward the listening party and back to the speaker every seven seconds. When turning toward the listening party, the robot makes a sad face. [N3] Attend to the other party when speaker changes. [N4] Turn toward the listening party and back to the speaker every seven seconds. When turning toward the listening party, the robot makes a sad face.}

Algorithm 4.10: This algorithm selects a specific robot response (manifestation) when the insensitivity state is active.

Intervention Selection Algorithm for Encourage Empathy Behavior

```
//The robot has six possible manifestations of the encourage emapthy behavior.
//We set a bool corresponding to each to false because they have not been enacted.
1. enacted_attend_to_each = false
2. enacted_ahem_attend_to_each = false
3. enacted_verbal_orient_to_each_other = false
4. enacted_verbal_amplify_negative_affect = false
5. enacted_nonverbal_amplify_negative_affect = false
6. enacted_ahem_nonverbal_amplify_negative_affect = false
//Check to see if the state is negativity state is active. If the negativity state is
//active, then trigger the intervention. If it is not active, wait and check again.
7.  While TRUE
8.          Initialize x_ins  //Indicator of whether or not insensitivty is active
9.          x_ins = state_var->safe_read() //Thread safe read (function g sets value)
            //If the state is active trigger the next intervention. These
            // interventions are shown in Algorithms 4.11-4.14.
10.         If x_ins == active
11.             If  enacted_attend_to_each == false
12.                 do_attend_to_each() //See algorithm 4.11
13.                 enacted_attend_to_each  = true
14.             Else If  enacted_ahem_attend_to_each == false
15.                 do_ahem_attend_to_each() //See algorithm 4.11
16.                 enacted_ahem_attend_to_each  = true
17.             Else If  enacted_verbal_orient_to_each_other == false
18.                 do_verbal_orient_to_each_other() //See algorithm 4.12
19.                 enacted_verbal_orient_to_each_other = true
20.             Else If  enacted_verbal_amplify_negative_affect == false
21.                 do_verbal_amplify_negative_affect() //See algorithm 4.13
22.                 enacted_verbal_amplify_negative_affect = true
23.             Else If  enacted_nonverbal_amplify_negative_affect == false
24.                 do_nonverbal_amplify_negative_affect() //See algorithm 4.14
25.                 enacted_nonverbal_amplify_negative_affect = true
26.             Else If  enacted_ahem_nonverbal_amplify_negative_affect == false
27.                 do_ahem_nonverbal_amplify_negative_affect() //See alg. 4.14
28.                 enacted_ahem_nonverbal_amplify_negative_affect = true
29.             End If
30.             Sleep_for_seconds(60)
31.         Else
32.             Sleep_for_seconds(5)
33.         End If
34. End While
```

The first intervention had the robot attend to the speaking dyad member while she spoke. Every seven seconds the agent would turn to the other person in the discussion quickly and then turn back to the speaker. The robot did this to orient the dyad members to each other. The response was meant to encourage the dyad members to recognize that they were involved in a discussion, and they needed to lend each other attention. The second intervention mimicked the first but with the "ahem" placed in front to draw attention to the agent. The implementation of these interventions appears in Algorithm 4.11.

Algorithm 4.11: This algorithm implements the attend to the other intervention. This is the first intervention when the insensitivity state is active. The second intervention is the same aside from the robot making an "ahem" sound to draw attention to itself.

Algorithm – do_attend_to_each()

```
//Attend to the party that is speaking. Periodically turn to the other party. Repeat
//with the second party when she begins speaking. This is to encourage each dyad
//member to remember the other party and view the disagreement as a discussion
//rather than two people simply making points at each other.
If  lower_power_member_is_most_recent_speaker
        turn_head_to_lower() //Turn toward the lower-power member
Else
        turn_head_to_higher() //Turn toward the higher-power member
End If
While most_recent_speaker_is_same
      wait_for_seconds(7)
      turn_toward_the_other_dyad_member_and_back()
End While
turn_head_180() //Attend to the other dyad member
While most_recent_speaker_is_same
      wait_for_seconds(7)
      turn_toward_the_other_dyad_member_and_back()
End While
reset_robot_to_default_state()
```

In the third intervention, the robot waited until a break in the conversation before saying that it was enjoying the chat between the two dyad members ("I'm enjoying the chat you two are having") and smiling. This, again, was to frame the discussion as ongoing dialogue between the dyad members so that neither dyad member ignored the presence of the other. This intervention's implementation is shown in Algorithm 4.12.

Algorithm 4.12: This algorithm implements the verbally orient to each other intervention. This is the third intervention when the insensitivity state is active.

Algorithm – do_verbal_orient_to_each_other()

```
//The agent waits until neither member of the conversation is speaking and says
//that it is enjoying the chat the two are having. This is to remind the dyad that it
//is a discussion between the two of them and to listen to each other's ideas.
1. Wait_until_pause_in_conversation()
2. Speak("I'm enjoying the chat you two are having.")
3. smile()
4. wait_seconds(5)
5. reset_robot_to_default_state()
```

Subsequently the agent tried to amplify the emotion in the interaction so that the higher-power dyad member had the opportunity to hear the emotion being projected by the lower-power dyad member from a neutral source. In the fourth intervention, the robot said that it sensed some discomfort during the discussion and frowned (again it asked if this assessment was correct to not appear commanding). The frown was used to associate the robot frowning with discomfort during the discussion (not an individual's emotion). The implementation of this intervention is shown in Algorithm 4.13.

Algorithm 4.13: This algorithm implements the verbally amplify the negative affect intervention. This is the fourth intervention when the insensitivity state is active.

Algorithm − do_verbal_amplify_negative_affect()

```
//The agent waits until neither member of the conversation is speaking and says
//that it senses discomfort in the discussion. This is to amplify the emotion in the
//discussion so that the dyad members can recognize its presence and confront it
//directly.
1. Wait_until_pause_in_conversation()
2. Speak(" I've sensed some discomfort during the discussion. Is that correct?")
3. display_sad_face()
4. wait_seconds(5)
5. reset_robot_to_default_state()
```

The fifth and sixth interventions saw the robot attend between the speakers, while checking in with the listening party every seven seconds (similar to the first two interventions). While checking in with the listening party, the robot frowned to amplify the discomfort in the discussion. These final interventions tried to amplify the fact that it was a discussion between the two parties and amplify the emotion in the room. The sixth intervention again contained the verbal "ahem" before it's start. The implementation of these interventions is shown in Algorithm 4.14.

When the acceptable relationship state is active, the robot wants to maintain this state. In this agent's implementation, the maintain behavior simply has the robot continue with its default behavior. This default behavior is an unengaged behavior (Table 4.12). The robot simply moves randomly to appear animated, but the movements have nothing to do with content of the conversation or the behaviors or the emotions of the dyad members. The robot engages in this behavior for the entirety of the interaction between dyad members in the baseline study described in Chapter 5.

Algorithm 4.14: This algorithm implements the attend to the other intervention. This is the first intervention when the insensitivity state is active. The second intervention is the same aside from the robot making an "ahem" sound to draw attention to itself.

Algorithm – do_nonverbal_amplify_negative_affect()

```
//Attend to the party that is speaking. Periodically turn to the other party. Repeat
//with the second party when she begins speaking. This is to encourage each dyad
//member to remember the other party and view the disagreement as a discussion
//rather than two people simply making points at each other.
If lower_power_member_is_most_recent_speaker
        turn_head_to_lower() //Turn toward the lower-power member
Else
        turn_head_to_higher() //Turn toward the higher-power member
End If
While most_recent_speaker_is_same
      wait_for_seconds(7)
      display_sad_face()
      turn_toward_the_other_dyad_member_and_back()
      reset_face_to_neutral()
End While
turn_head_180() //Attend to the other dyad member
While most_recent_speaker_is_same
      wait_for_seconds(7)
      display_sad_face()
      turn_toward_the_other_dyad_member_and_back()
      reset_face_to_neutral()
End While
reset_robot_to_default_state()
```

Table 4.12: The Implemented Intervention for the Maintain Behavior

$\beta_{maintain,1}$ {<Nonverbal cues> [Nonverbal Cue Order] [N1] Unengaged Behavior (Random arms swings, head nods, head turns, hip twists, and blinks.)}

This subsection provided a description of the interventions for the autonomous robotic agent as well as how these interventions are chosen. The interventions are rooted in the practices and guidelines of transformative mediation. The successes and shortcomings of these interventions help to answer the third subsidiary question from Chapter 1. The third question asks what channels of communication a robotic agent should use and how overt the agent should be when intervening to support positive change in a hierarchical human-human relationship. These interventions have been incorporated into the in the autonomous agent tested in the study described in Chapter 6.

4.4 Summary

This chapter operationalized the states of the computational model introduced in Chapter 3. The problematic relationship states are states of weakness and alienation that are identified by transformative mediators in relationship-focused mediation literature. The chapter described how these states are rooted in the transformative mediation literature and provided rating scales so human raters could identify them. It then provided an implementation for an autonomous robot to identify two states and respond to these two states with interventions that followed the guidelines followed by professional mediators.

The autonomous robot developed as part of this thesis used the implementations described here in the study described in Chapter 6. The following chapter describes a baseline study in which the data to inform this implementation was gathered. This baseline study employed a robot that remained unengaged throughout the interaction by demonstrating the behavior $\beta_{maintain,1}$ mentioned above. This baseline provides a basis to which the study conducted with autonomous robot can be compared and helps to answer questions about how the mere presence of a robot will influence strained hierarchical relationships in conflict.

CHAPTER 5

BASELINE HUMAN-ROBOT INTERACTION STUDY

To say that a robot can successfully support positive transitions in hierarchical human-human relationships embroiled in negative conflict, one must be able to compare results gathered when a robot acts to support the relationships to results when the robot does not act to support these relationships. The study presented in this chapter is designed to create a conflict with negative dynamics in a hierarchical relationship. During this conflict, a robot that has been introduced as something to help with communication is present but does not intervene. Instead, this unengaged robot simply moves in subtle ways to give the impression it may intervene. The results reported in this chapter were first reported in Pettinati and Arkin (2019). The relationship between this work and the subsidiary questions asked in chapter one is summarized in Table 5.1.

The first five sections of this chapter provide the details of the study's setup and execution. The sixth and seventh sections provide the results of the study and discuss what these results have to say about the first, third, and fourth subsidiary questions from chapter one. The seventh section of the chapter discusses the training of the autonomous system to identify states of negativity and insensitivity in the relationships. The insights gleaned from the training process are discussed with respect to how they help to answer questions two and three from chapter one. The chapter concludes with a discussion section.

Table 5.1: Each subsidiary question from chapter one is presented along with how this experiment helps to provide an answer.

Relationship Between Experiment and Subsidiary Questions	
Subsidiary Question	Relationship to Experiment
How is an unengaged robot, that has been introduced as something to support communication in the relationship, perceived by the dyad members and how do interventions change this perception?	• Self-report and objective data are collected about how the participants perceive and treat the robot. • These data can then be compared to data collected with an intervening robot to assess how interventions change these perceptions.
How can a social robotic agent represent a dyad's relationship state, problematic or otherwise, and decide when to intervene in the relationship?	• The experiment provides a small corpus of examples of the problematic states defined in the previous two chapters, the insensitivity and negativity states. • These examples allowed for us to train an autonomous system to identify these problematic relationship states. • The work to train this system reveals important markers of these states for autonomous systems as well as potential future work.
What channels of communication should the agent use when supporting the dyad, how overt should these communications be in order to avoid upsetting the relationship, and how should the agent choose between interventions?	• The accuracy with which problematic relationship states can be automatically identified can influence the ways in which the robot intervenes. • The data collected with an unengaged robot can be compared to data collected with an intervening agent to understand how the interventions changed relationship dynamics (beyond mere presence effects).
What issues does the power dynamic in the relationship present for the robotic agent when trying to provide a conflict process that is viewed as fair and equitable by both participants?	• The robot may be perceived by or affect the behavior of lower-power and higher power dyad members differently. • The way in which the robot intervenes could also influence the perceptions of whether or not the robot is fair. • This study provides baseline results for the how the robot is perceived by higher and lower-power dyad members.

5.1 Study Overview

This study has two between-subject conditions. In both conditions, there is an unengaged robot present. This means there is a robot that moves subtly and in ways that are not correlated with the conversation that is taking place. This behavior is further described in section 5.3. The robot used in the study is Robokind's R25[11] shown in Figure 5.1. In one of the two conditions, a participant is assigned a lower-power role in a discussion with an experiment confederate where there is conflict. In the other condition, the participant is assigned a higher-power role in the discussion with the experiment confederate. The way in which the higher and lower-power designation is made is described in this section just below.

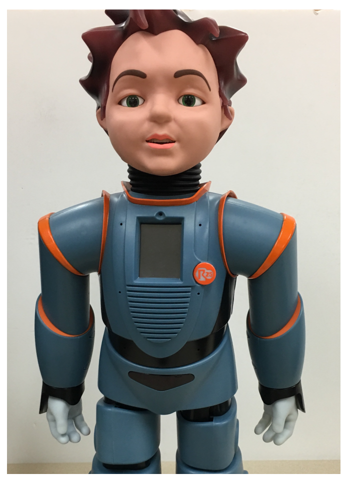


Figure 5.1: This is Robokind's R25. This robot was used for all of the study's conditions. It has the ability to display affect through multiple nonverbal channels, facial and bodily kinesics.

[11] https://www.robokind.com/robots4autism/meet-milo

During the experiment, a confederate and a participant are generating an argument about a recent law change in Georgia that divided the student population on the Georgia Tech campus. Specifically, they construct an argument where they choose and defend one side of the issue: "I feel less safe on the Georgia Tech campus with it being legal for permit holders to conceal carry a gun, and it should not be allowed". We chose this topic by surveying a class of Georgia Tech students on several controversial topics. This is the topic that most divided the class.

The confederate and participant are told that they have fifteen minutes to discuss the issue and generate a single argument before the experimenter enters back into the room to hear a final argument. They are told that one member of the pair will make this final argument and that the final argument will determine compensation for both of them. The compensation is allegedly assessed based on provided guidelines (Figure 5.2 on the next page).

The higher-power member in the relationship is the member who makes the final argument. This relationship member has direct control over the compensation; she is explicitly told that she can take the fifteen-minute discussion into account when arguing or not. The higher-power member's only responsibility is to make the strongest possible final argument. The lower-power individual is told to use the fifteen minutes to present her ideas, but she will have to remain silent during the final argument. The responsibility of the lower-power member is to help the make the final argument as strong as possible. Half of the participants are assigned the higher-power role, and half are assigned the lower-power role.

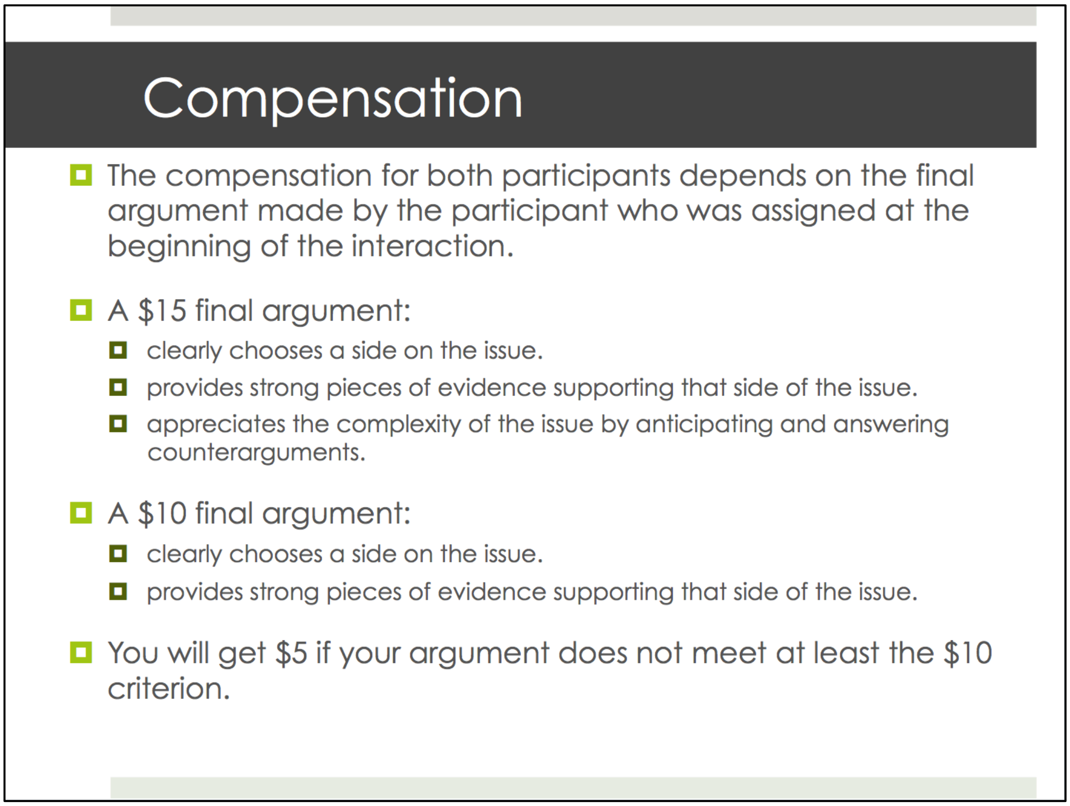


Figure 5.2: The compensation guidelines sheet provided to participants during the setup of the discussion. Participants are told that a two-minute final argument determines compensation for both of them. The higher-power individual makes the final argument. The lower-power individual does not contribute directly to the final argument. The compensation is determined based on the guidelines given in this sheet.

For clarity in this discussion, let us assume that the confederate is a male (which he was) and the subject is female (participants from both genders were included in this study). The confederate, regardless of his role, helps to drive conflict, using provided guidelines, such that the participant is put in a weakened state and becomes alienated from him early in the interaction.

After the participant has the opportunity to make a point for her side of the issue, the confederate uses the language identified in the transformative mediation literature (e.g. Moen et al., 2001) that shows a refusal to recognize the other party's thoughts and feelings. The confederate will use the language from the literature that shows he views the participant's thoughts and feelings about the topic at hand as invalid and unworthy of consideration for the argument (e.g. "Okay, well to be honest, I can't understand how you can say that…").

After dismissing the participant's viewpoint, the confederate argues his side with language from the literature that shows weakness (uncertainty and/or frustration) (e.g. "I'm just not sure…" or "I guess"). The participant will see someone who does not seem to respect her thoughts and seems unable to change his mind. Alienation from the confederate drives weakness in the participant (Bush & Folger, 2010a).

Finally, the confederate makes a remark that requests recognition from the participant (e.g. "It's important to see…" "You have to understand that…" or "The reason I am saying this is…"). The confederate requests recognition (using the language of Moen et al., 2001) at the conclusion of each of his arguments. These statements say, "I want you to understand me", and require acknowledgment of the legitimacy of the confederate's feelings. These statements allow for the participant to dictate the tone of the interaction.

After the confederate drives conflict early in the interaction, the confederate mirrors the responsiveness of the participant. If the participant does not offer recognition, weakness continues. The confederate continues using responses that follow the pattern described above. If the participant gives recognition, the confederate responds with strength and recognition.

Showing strength means speaking clearly and without the language of frustration or uncertainty. Showing recognition means being attentive and open to the participant when the participant is speaking. It also means acknowledging the validity of the other's feelings and respecting the other's thoughts about the topic (e.g. Bush & Folger, 2010a). The confederate's new pattern (until the participant refuses recognition again) is showing emerging understanding, making a point on his side of the issue using stronger language, and making another request for recognition. To show recognition, the confederate explicitly says something that shows he is considering the participant's viewpoint, and he is willing to engage the participant directly even if he disagrees with the participant's viewpoint. The confederate decides whether or not to respond with strength and recognition or weakness and alienation based on whether the participant gives recognition.

Recognition is given if: the participant explicitly acknowledges the validity of the argument the confederate has just made, explicitly says he has made a point that is important for the argument, considers the confederate's argument in more detail, e.g. by saying she would appreciate that argument more "if…" or by asking questions regarding the argument, or if the participant admits that she is thinking about the topic in a new way. The participant dictates whether or not mutual recognition is given in the interaction. A participant's willingness to give recognition to the other depends in part on his/her strength/degree of empowerment (e.g. Bush & Folger, 2010a).

5.2 Participants

There were a total of 31 participants tested during this study. The participants were recruited via flyers hung around the Georgia Tech campus as well as campus mailing lists. The emails to these mailing lists contained exactly the same call as the flyers. The flyer used for advertising is included as part of Appendix D.

The participants were all Georgia Tech students who confirmed that they held a strong opinion on the conceal carry on Georgia Tech issue (introduced above). We restricted the study to this population because the topic is meaningful to Georgia Tech students and, perhaps, is less meaningful to those outside of the Georgia Tech community. People from outside the campus may not have these same strong feelings about the issue because they are not regularly attending events on campus. We needed participants who care about and want to be heard on this issue; otherwise, there may not have been emotional investment. If participants were not emotionally invested, then the dynamics we're studying would not emerge.

During the 31 trials, there was 1 trial where the robot malfunctioned and did not move at all, and there were 2 participants who claimed that they knew the confederate was a member of the study team during the discussion. These 3 trials were excluded from analysis. Within the 28 remaining trials, there were 14 participants who had the lower-power role in the relationship, and there were 14 participants who had the higher-power role in the relationship. The demographics information (including: the genders, ages, opinions about concealed carry, and backgrounds in artificial intelligence of these participants) is summarized in Tables 5.2-5.6 and Figures 5.3-5.7 for both conditions. All of this information was gathered immediately after the consent form was signed (before the participants interacted with the robot.

Table 5.2: Both conditions were predominately male. There were a total of 21 males that participated in the study compared to only 7 females.

Breakdown of Participant Gender			
Condition	Male	Female	Nonbinary
Lower-Power Participant	11	3	0
Higher-Power Participant	10	4	0

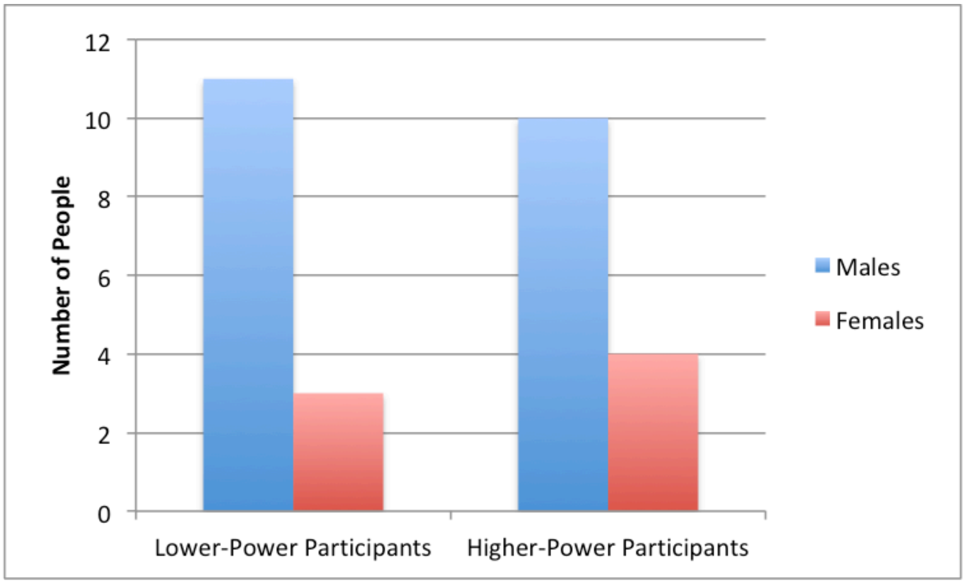


Figure 5.3: In both conditions, the number of male participants outnumbered female participants (11 to 3 and 10 to 4).

Table 5.3: The participants' ages ranged from 18 to 32. Participants in both groups were mostly undergraduates at Georgia Tech, so the average age for both groups fell around 21.

Breakdown of Participant Age		
Condition	Mean Age	Standard Deviation
Lower-Power Participant (n = 14)	21.57	4.309
Higher-Power Participant (n = 14)	20.93	2.556

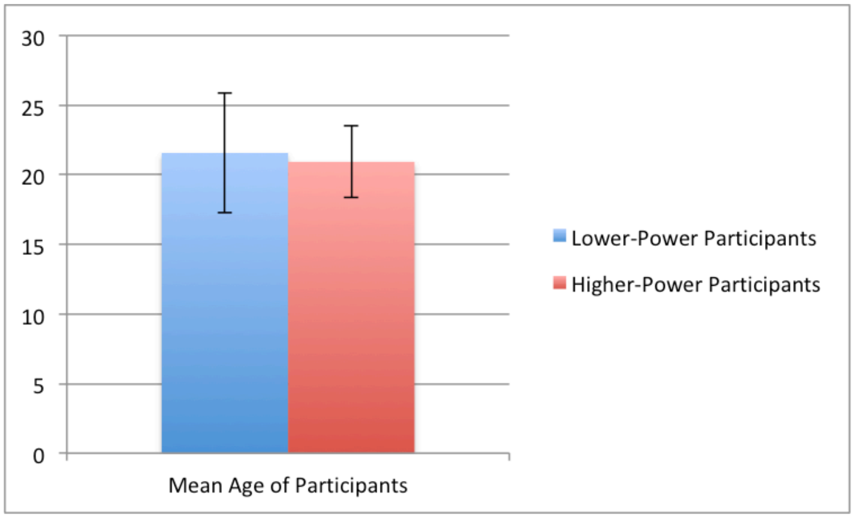


Figure 5.4: The ages of participants for the participants for both groups was the early 20s. There was not a significant difference between groups with respect to age.

Table 5.4: In both groups, most participants were against the passage of the new law dealing with the concealed carry. There were 24 people who entered the conversation believing that the concealed carry of a gun should be banned on the GT campus compared to only 4 participants who supported the law change.

Number of Participants on Each Side of the Argument		
Condition	Against Concealed Carry	For Concealed Carry
Unengaged Robot, Lower-Power Participant	12	2
Unengaged Robot, Higher-Power Participant	12	2

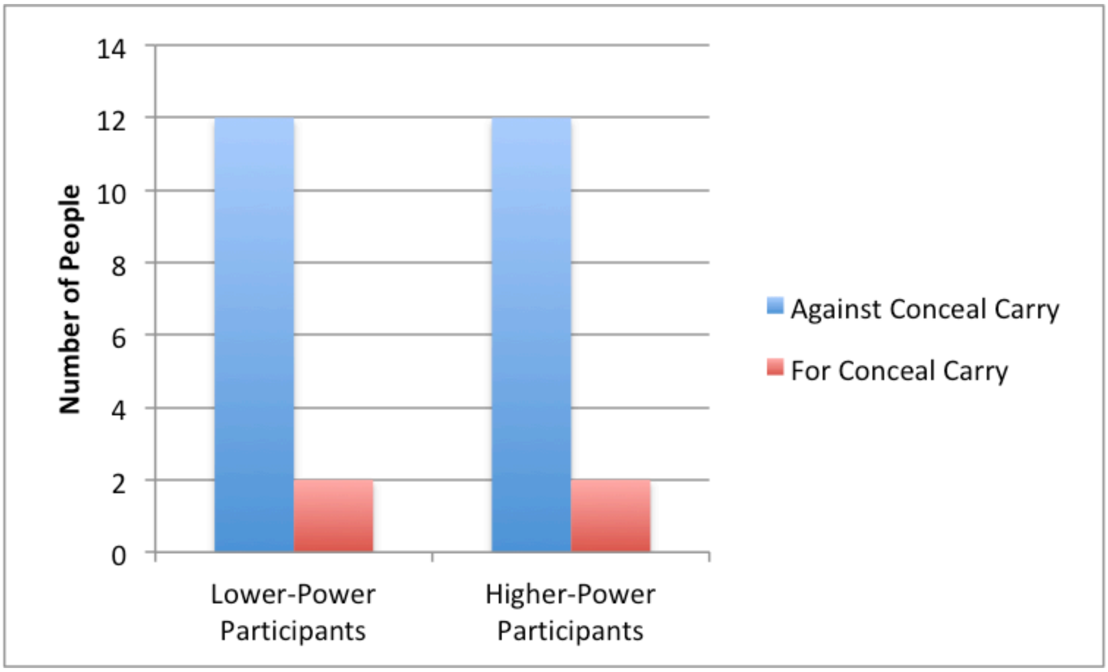


Figure 5.5: There were only two participants in each condition who felt as safe or safer with the concealed carry law change. There were twelve participants in each condition that felt less safe with the conceal carry of a gun being allowed on Georgia Tech.

Table 5.5: There were very few participants (1 in each conditions) that did not pay attention to current events. Most participants paid attention to some degree but few engaged others about what was happening.

Summary of Participants' Media Consumption Habits				
Condition	I don't keep a close eye on current events.	I watch or read the news once in a while.	I watch or read from a variety of news sources regularly.	I watch or read from a variety of news sources regularly and actively engage others in discussion about various issues.
Lower-Power Participant	1	6	5	2
Higher-Power Participant	1	5	6	2

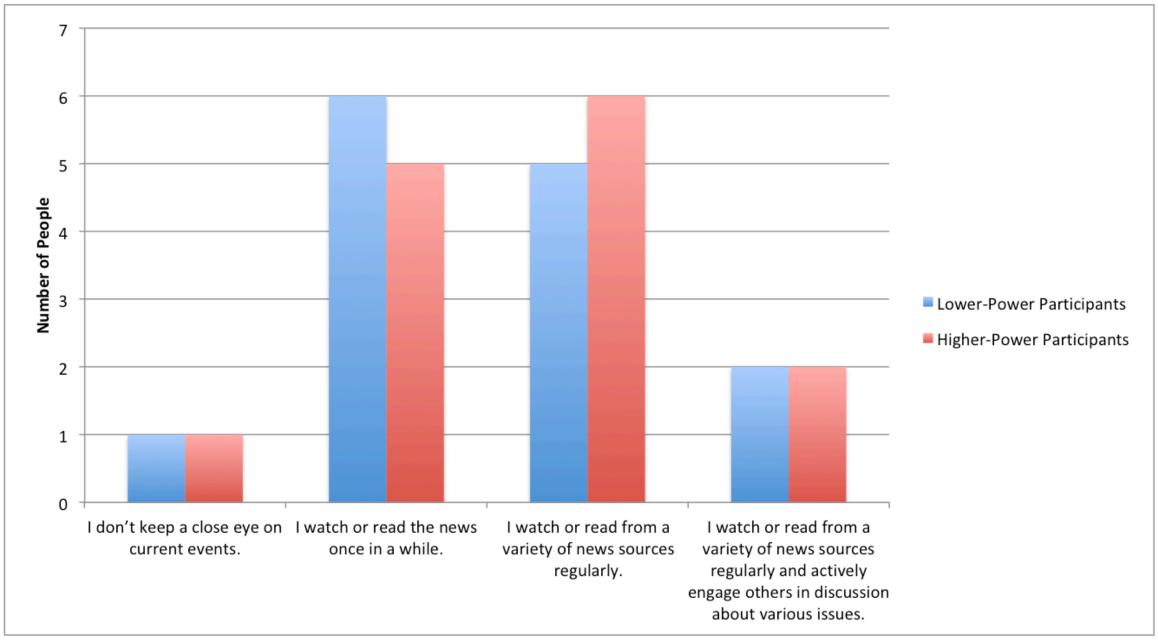


Figure 5.6: There was a unimodal distribution in both groups when it came to their consumption of current events. In both groups, participants paid attention to the news to some degree, but few participants regularly engaged others in discussion about what was going on.

Table 5.6: Most participants had little experience with artificial intelligence. 26 of the 28 participants did not have any formal artificial intelligence background.

Summary of Participants' Backgrounds in AI				
Condition	Exposed to AI in Pop Culture	Had Done Casual Independent Study of AI	Had Taken University Level Classes in AI	AI Researcher
Lower-Power Participant	8	6	0	0
Higher-Power Participant	9	3	2	0

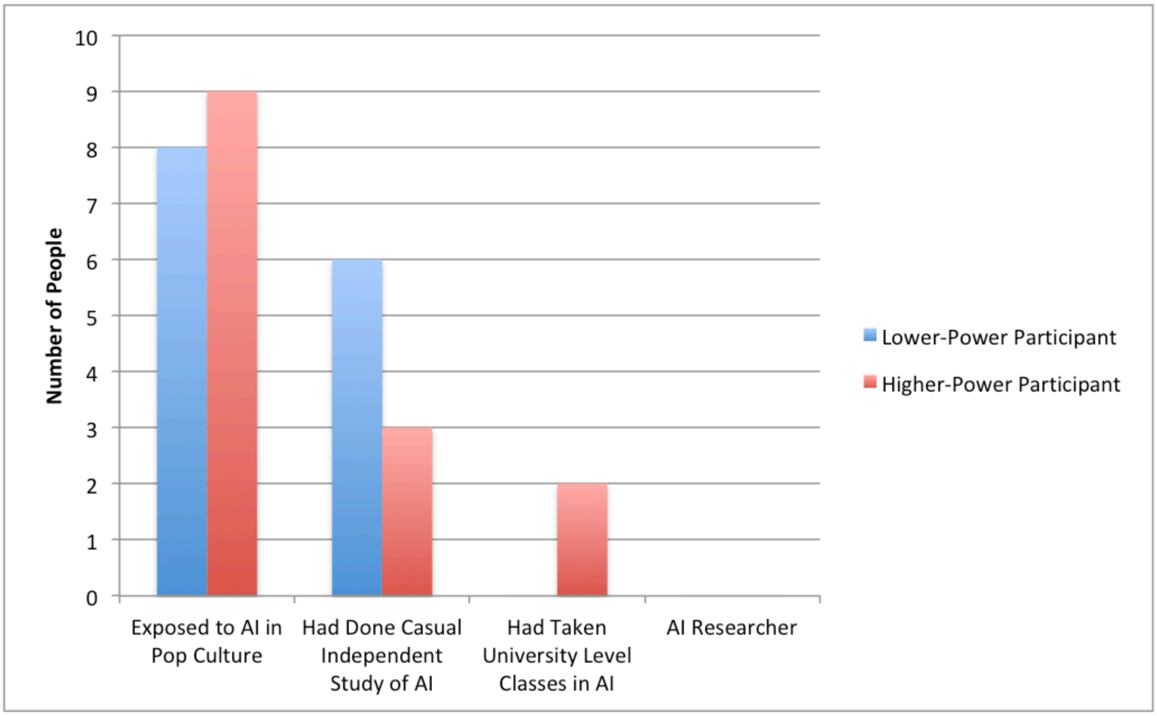


Figure 5.7: The majority of participants in both conditions had only been exposed to artificial intelligence in popular culture. None of the participants in either condition had advanced knowledge on the topic.

Participants completed self-report measures (Likert-style measures ranging from 1 to 7) at the outset of the study that related to their perceived dominance, their views on interpersonal conflict, and their views on robots helping in interpersonal relationships. These measures are included as part of Appendix D. These results are summarized in Tables 5.7-5.10 and Figures 5.8-5.11.

Participants were neutral in their willingness to lead discussions and their tendency to simply accept the ideas expressed by others. They claimed to be comfortable in expressing their ideas on contentious topics and confident in how they expressed themselves (Table 5.7 and Figure 5.8). With regard to interpersonal conflict, they were neutral when asked about whether or not in hierarchical relationships the needs and wants of the higher-power individual are often put above those of the lower-power individual (Table 5.8 and Figure 5.9). They agreed that often two people in conflict often need a third-party to help resolve the conflict (Table 5.9 and Figure 5.10).

Finally, participants seemed open, although reserved, about the possibility of a robot helping relationships in conflict. They were neutral about a robot making them more uncomfortable than a third person during a dispute, and they were neutral about a robot being able to understand and reflect how they were feeling. Participants slightly disagreed that a robot could not do anything to help human-human relationships, and they slightly disagreed that they would feel judged if a robot tried to reflect their emotions (Table 5.10 and Figure 5.11).

Table 5.7: The participants completed Likert-style scales ranging from 1 (strongly disagree) to 7 (strongly agree). The value 4 was uncertain. The means and standard deviations for each statement are shown. The participant was unsure about leading group discussions and going along with others' ideas. Participants were confident expressing themselves and expressing controversial opinions. These measures were gathered before the study's discussion.

Summary of Participants' Trait Dominance In Both Groups		
Condition	Mean	Standard Deviation
	Responses to: I like to lead when having a group discussion.	
Low Power Participant (n = 14)	4.57	1.4
High Power Participant (n = 14)	4.5	1.16
	Responses to: I go along with other people's ideas when deciding what to do.	
Low Power Participant (n = 14)	3.92	1.14
High Power Participant (n = 14)	4.14	1.41
	Responses to: I am comfortable expressing my opinions on a contentious topic.	
Low Power Participant (n = 14)	5.643	0.63
High Power Participant (n = 14)	5.286	1.49
	Responses to: I am usually confident in how I express my ideas and myself.	
Low Power Participant (n = 14)	5.92	0.92
High Power Participant (n = 14)	5.36	1.08

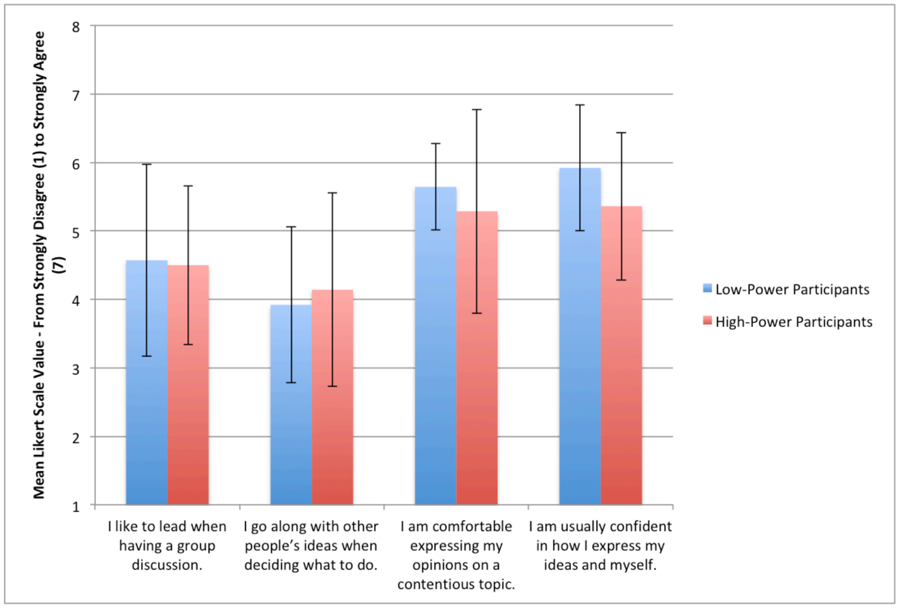


Figure 5.8: There were no differences between groups on measures related to participant dominance during group discussions.

Table 5.8: Participants were unsure (4) if the needs of lower-power individuals were often overlooked in hierarchical relationships. They were also unsure (4) about their ability to see other perspectives when having a disagreement.

Summary of Participants' Views on Hierarchical Human-Human Conflict		
Condition	Mean	Standard Deviation
	Responses to: When I disagree with a superior, I often feel obligated to put their needs and wants above my own.	
Low Power Participant (n = 14)	4.0	1.24
High Power Participant (n = 14)	4.64	1.28
	Responses to: When I openly disagree with someone who has more control than I do in a situation, I often feel like my ideas aren't seriously considered.	
Low Power Participant (n = 14)	4.429	1.34
High Power Participant (n = 14)	4.857	1.17
	Responses to: When I have strong feelings about a topic or decision, it can be difficult to listen to and appreciate alternative viewpoints.	
Low Power Participant (n = 14)	4.071	1.54
High Power Participant (n = 14)	3.357	1.5

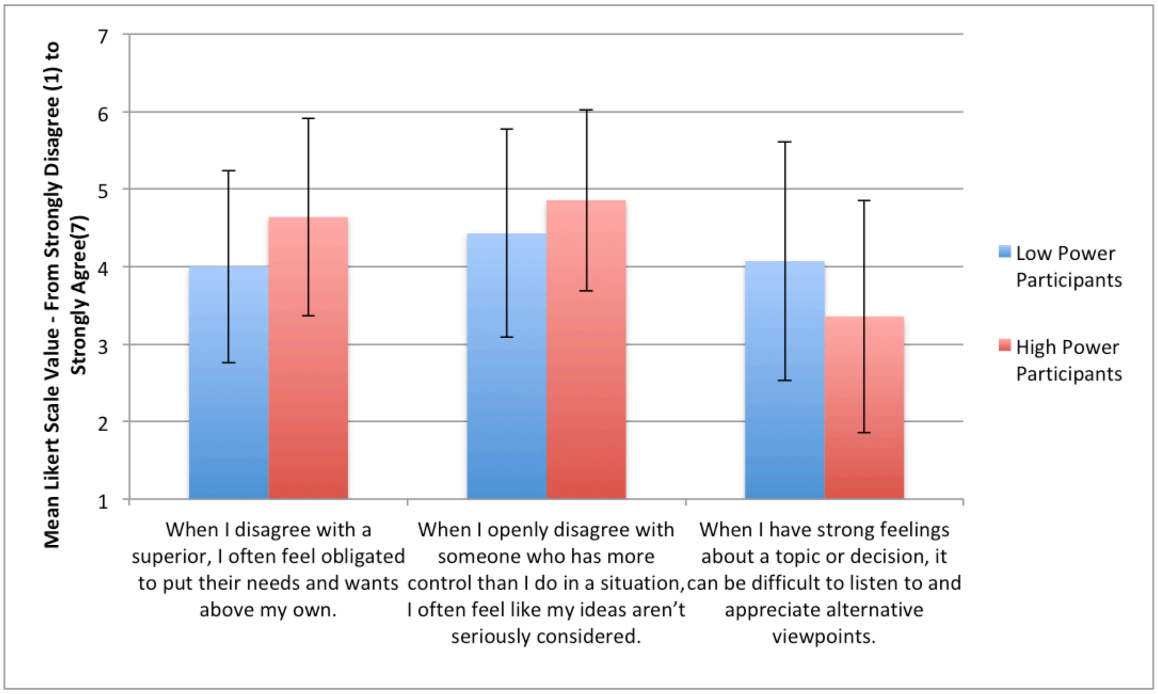


Figure 5.9: There were no significant differences between groups on how participants perceived conflict in hierarchical relationships.

Table 5.9: Participants in both groups agreed (slightly agree = 5) that a dyad sometimes needs a third-party to overcome issues that they are having.

Condition	Mean	Standard Deviation
	Response to: Sometimes two people who disagree need a neutral third-party to help them listen to one another and come to mutually satisfactory resolution.	
Unengaged Robot, Low Power Participant (n = 14)	5.857	0.95
Unengaged Robot, High Power Participant (n = 14)	5.42	1.09

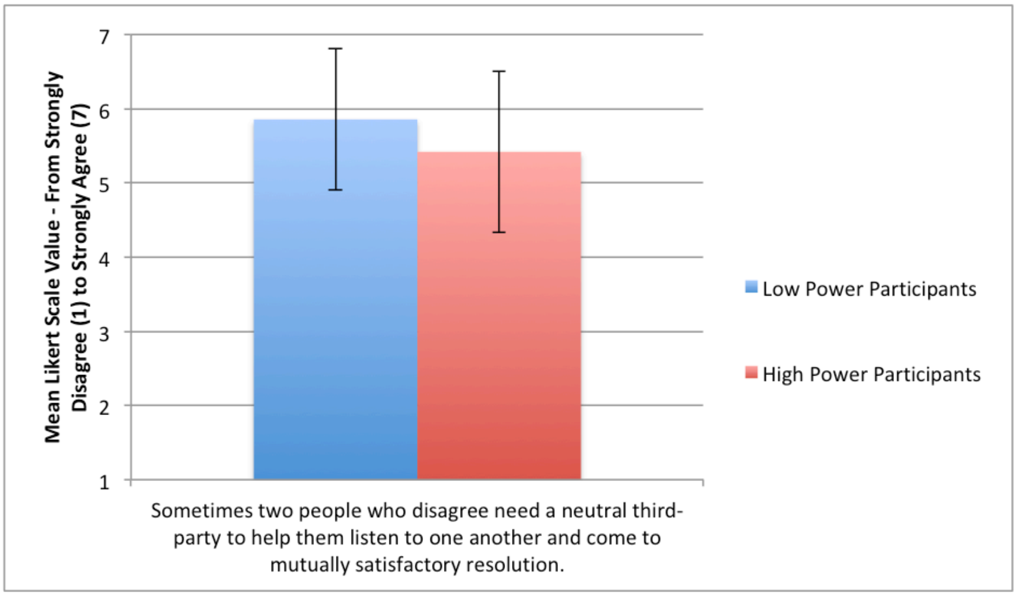


Figure 5.10: The participants in both groups believed that a third-party can sometimes be necessary for conflict resolution. The value of 5 indicated that the participants slightly agreed with the statement. There was not a significant difference between groups.

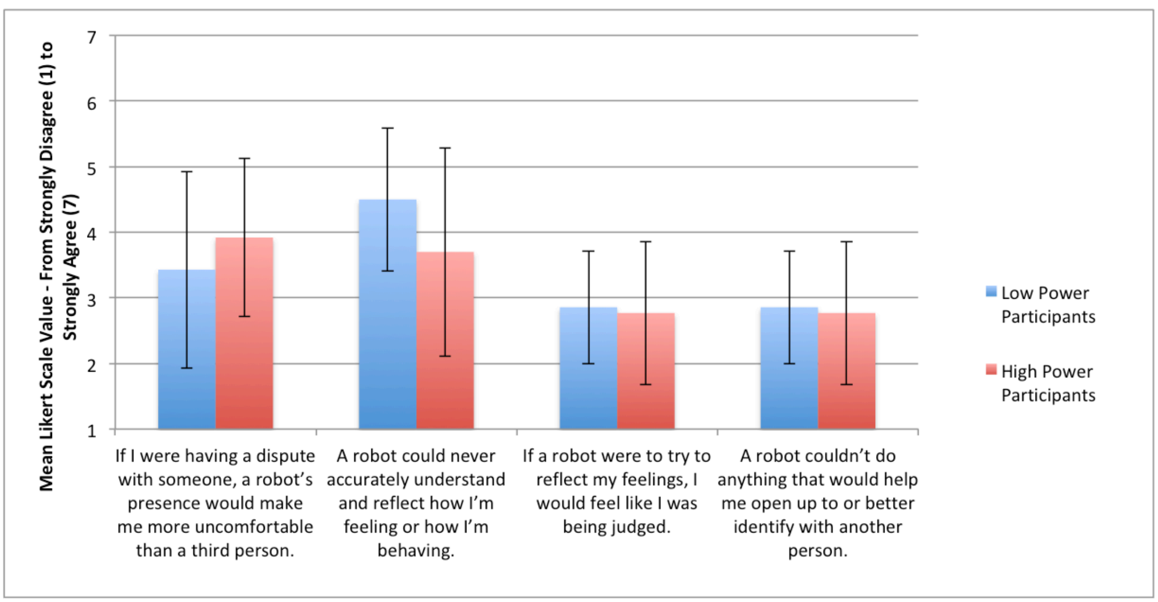


Figure 5.11: Most participants were unsure (4) about how the presence of a robot would make them feel during conflict. Participants seemed to be open to the idea of a robot helping with human-human interaction; they slightly disagreed (3) with the statement that robots couldn't do anything to help two humans identify with one another.

Table 5.10: Results for Likert-style measures ranging from 1 (strongly disagree) to 7 (strongly agree) related to how participants felt about a robot trying to support human-human discussions. Participants were unsure about how the presence of a third-party robot would make them feel as well as whether or not a robot could understand their feelings. They slightly disagreed that they would feel judged by a robot or that a robot could not help to support human-human interactions.

Participants' Impressions of Robots Entering the Discussion		
Condition	Mean	Standard Deviation
	Responses to: If I were having a dispute with someone, a robot's presence would make me more uncomfortable than a third person.	
Low Power Participant (n = 14)	3.43	1.5
High Power Participant (n = 14)	3.92	1.21
	Responses to: A robot could never accurately understand and reflect how I'm feeling or how I'm behaving.	
Low Power Participant (n = 14)	4.5	1.09
High Power Participant (n = 14)	3.7	1.59
	Responses to: If a robot were to try to reflect my feelings, I would feel like I was being judged.	
Low Power Participant (n = 14)	2.857	0.86
High Power Participant (n = 14)	2.769	1.09
	Responses to: A robot couldn't do anything that would help me open up to or better identify with another person.	
Low Power Participant (n = 14)	2.857	0.86
High Power Participant (n = 14)	2.769	1.09

5.3 Unengaged Robot Behavior

During the experiments, in both conditions, the robot maintained an unengaged behavior for the entire fifteen-minute conversation. The unengaged robot makes certain movements periodically. The periods of the movements are long to prevent the participant from recognizing that the robot is repeating movements at set time intervals. The robot will blink, twist its head slightly and slowly side to side, tip its head up and down slightly and slowly, swing one or both arms, and stretch its hips slightly. These movements are very subtle and slow. They intentionally show no indication of the robot following the discussion or attending to anything about either the confederate or the participant. They were programmed to make the robot not stationary (frozen) during the interaction.

5.4 Study Procedure

The step-by-step procedure for the experiment is given in Table 5.11. The experiment took place in Georgia Tech's Tech Square Research Building (TSRB). The experimenter greeted the confederate and participant in the lobby of this building before bringing them to the second floor. The discussion on concealed carry on Georgia Tech took place in an office on the second floor of TSRB (shown in Figure 5.12). The forms for the study were completed at cleared off desks just outside of the office.

Figure 5.12: This is a picture of the office as it organized during an experiment trial. The participant and confederate face each other directly. The robot stands at the periphery of the interaction as a bystander.

Table 5.11: The step-by-step procedure for the experiment.

Step 1: The experimenter greets the participant and confederate in building lobby	• At the time the participant is scheduled to arrive, the confederate leaves through the back of the research building and enters through the front doors to appear to be just another participant in the study. • The experimenter goes directly to the lobby to greet the confederate and the participant. • After greeting both the participant and the confederate, the experimenter leads them upstairs to the second floor of the building.
Step 2: The confederate and participant complete the consent forms and demographics and personality survey	• The consent form and demographics/personality survey are completed at desks just outside the office in which the discussion on concealed carry takes place. • The experimenter gives the confederate and participant the consent form, the video release form, and the demographics survey one at a time. • The confederate pretends to fill out these forms until the participant is done with each one.
Step 3: The experimenter affixes sensors to the participant and confederate in the office.	• The participant and confederate are led into the office where the discussion takes place and asked to have a seat in their respective locations (Fig. 5.12). • They have identical sensors affixed, a lapel microphone[12] as well as the Empatica E4 Wristband[13]. • They are told to take three minutes to clear their heads (without interacting) to relax before the discussion.

Step 4: The dyad members are given their objective for the discussion.	• Once the three-minute time period ends, the experimenter returns to the room.
	• The experimenter tells the dyad:
	  o To work together to form an argument for or against the following statement: "I feel less safe on the Georgia Tech campus with it being legal for permit holders to conceal carry a gun, and it should not be allowed"
	  o They have fifteen minutes to form the best single argument they can and the strength of that argument would be assessed using the guidelines shown in Figure 5.2.

[12] https://www.amazon.com/gp/product/B01AG56HYQ/ref=ppx_yo_dt_b_asin_title_o05_s00?ie=UTF8&psc=1
[13] https://www.empatica.com/research/e4/

Table 5.11 (cont.): The step-by-step procedure for the experiment

Step 4 (cont.): The dyad members are given their objective for the discussion.	• The experimenter also tells the dyad: o The strength of their argument will determine their compensation. • This act of deception encourages the participant to have a stake in making a strong argument and helps to create weakness in the participant when the confederate begins to drive the conflict.
Step 5: The experimenter attains the views of the participant.	• The experimenter asks the participant whether or not she feels less safe with concealed carry of a gun being allowed on the Georgia Tech campus. • Once she has given an answer, the confederate is asked. The confederate always chose the position opposite that of the participant.
Step 6: The experimenter explains the structure of the interaction.	• The dyad members are also told that the interaction is structured to allow for a fair and equitable discussion. o Each dyad member has one to two minutes to make a point and respond to what was previously said by his or her partner before he or she should give the other person another chance to speak.

Step 7: The dyad members are given their roles for the interaction (this sets up the relationship as hierarchical).	• The experimenter then assigns the participants their roles in the interaction, as described above, to establish the relationship hierarchy.
	• The higher-power member is explicitly told that:
	  o She/he will make the two-minute final argument that determines compensation for both members of the dyad
	  o She/he can take into account the 15-minute discussion or ignore what has been said during the discussion when making the final argument
	• The lower-power member is explicitly told that:
	  o She/he has no say in the final argument and must remain silent during the final argument
	  o She/he can only contribute during the 15-minute discussion

Step 8: The experimenter introduces the robotic agent.	• Finally, the robot is introduced to the dyad as something to support their communication. • They are told it knows nothing about the topic at hand, and it may or may not intervene in the discussion. • The robot begins the unengaged behavior described in Section 5.3. It maintains this for the entire study.

Table 5.11 (cont.): The step-by-step procedure for the experiment

Step 9: The confederate and the participant have the discussion on concealed carry.	• The experimenter then leaves the parties to have the discussion. The confederate follows the guidelines described above during the discussion. The confederate tries to appear just like another participant. • An extensive document that contains guidelines for the experimenter and the confederate is available on the Mobile Robot Lab website[14].
Step 10: The higher-power dyad member makes the final argument to the experimenter.	• After the fifteen minutes conclude, the experimenter enters the room and requests the final argument from the higher-power dyad member. • The higher-power dyad member gives the final argument, which is a maximum of two minutes. If the confederate is the higher-power dyad member, he uses a scripted argument. The lower-power dyad member is instructed to stay quiet during the final argument.
Step 11: The dyad completes the final measures and is debriefed.	• After the final argument is given, the experimenter tells the dyad that they will be compensated after they complete the final surveys. • The dyad is shown to desks outside of the office where they complete questionnaires. Once again the confederate pretends to be working until after the participant has finished, to avoid time pressure. • After the participant has completed the questionnaires, the participant is debriefed.

5.5 Measures

Objective, physiological, self-report, and video-rating measures were collected as part of this study. All of the measures are summarized in Table 5.12. Appendix E provides a more explicit overview of measure generation. The physiological measures were not used in the final analysis. Physiological signals have been shown to relate to psychological distress (e.g. Bradley & Lang, 2000). Certain participants in this study,

---

[14] https://www.cc.gatech.edu/ai/robot-lab/nri_thesis/thesis_supplemental/Guidelines.docx

however, came to the trial sweating, with a raised heart rate from running to the research building, etc. This made gathering accurate baseline readings very difficult, which limits one's ability to accurately identify problematic patterns in the data.

An experimenter generated the objective measures from the audiovisual recordings made during the study. This experimenter and a video coder generated the video-rating measures using the guidelines given in Appendix A. These instances of data generation introduce experimenter bias into outcome measures. Appendix E describes the outcome measure generation along with the considerations paid to experimenter bias.

Inter-rater reliability on the video-rating measures was assessed using Cohen's κ; a measure used in related works (e.g. Jung, 2016). The agreement between raters was $\kappa = .512$ on the weakness ratings. The agreement between raters was $\kappa = .502$ on recognition ratings. This is moderate agreement (Simand & Wright, 2005) and is comparable to the agreement by coders in previous studies (e.g. Jung, 2016).

Table 5.12: A summary of all the outcome measures used during the study

Measure Type	Individual Measures	Collection Method
Objective	• The participant's average speaking time (seconds) • The participant's total speaking time (seconds) • The length of the final argument by a higher-power participant (seconds) • The count of unique arguments presented by the participant during the discussion • The count of unique argument presented by the higher-power participant during the final argument • The number of times the participant mentioned his/her feelings (how he/she was presently feeling) • The number of times the robot was explicitly referenced by the participant • The participant's number gazes toward the robot	Coded by an experimenter watching audiovisual recordings
Physiological	• blood volume pulse • skin temperature • skin conductance	Empatica E4 wristband

Table 5.12 (cont.): A summary of all the outcome measures used during the study

Self-Report	• Perceived Emotional Conflict • Perceived Task-Related Conflict	Jehn's (1995;2001) intragroup conflict scales
	• Partner's perceived involvement and affection shown during the conversation • Partner's perceived receptivity to your ideas and trustworthiness • Partner's perceived depth/similarity/inclusion • Partner's perceived dominance • Partner's composure • Partner's formality • Partner's task vs. social orientation	Burgoon and Hale's (1984;1987) Relationship Communication Scale.
	• Participant's impressions of the robot ○ Whether or not it was a distraction ○ Whether or not it understood how he/she felt ○ Whether or not it was on his/her side	Likert-style items (on a scale of 1 to 5), See Appendix D
	• Participant's perceived level of empowerment and recognition	Likert-style items (on a scale of 1 to 5), Influenced by (Folger, 2010), See Appendix D
Video Rating	• Video Segments in which the participant was speaking or listening were labeled as showing a lack of empowerment (or emerging empowerment) ○ A participant's level of weakness was the proportion of video segments that showed a lack of empowerment to the total number of video segments coded for empowerment. • Video Segments in which the participant was speaking were labeled as showing a lack of recognition (or genuine recognition) ○ A participant's level of alienation was the proportion of video segments that showed a lack of recognition to the total number of video segments coded for recognition.	Coded by experimenter and video rater who was unaware of the experiment's hypotheses using the guidelines are included in Appendices A

5.6 Results

To compare the lower and higher-power groups on each of the measures introduced in the previous section, we used a two-sample (unpaired) two-tailed t-test. We used Levene's Test of Equal Variance (1960) to ensure that the groups did not violate the equal variance assumption. In the case they did, we ran a t-test where equal variance was not assumed. The t-test is the standard way of evaluating differences between groups in the literature and has been used in similar studies on measures we have incorporated here (e.g. Hoffman et al., 2015; Jung et al., 2015; Jung, 2016).

On certain measures, the normality assumption of the t-test did not hold. We used Shapiro and Wilk's (1965) test of normality to check the normality assumption for each group used in the t-test. When the data failed this test, we ran the Mann-Whitney U Test (a nonparametric test) (Mann & Whitney, 1947). Previous research has used nonparametric methods when the normality assumption of the t-test does not hold (e.g. Jung, 2016). To check the internal consistency of scales containing multiple items measuring a single construct, we used Cronbach's α (Cronbach, 1951). If the value fell below 0.7, we tested each item independently.

Jehn's (1995; 2001) emotional conflict scale had a Cronbach $\alpha = .381$. We ran three independent tests for each question. Jehn's task-related conflict scale had a Cronbach $\alpha = .888$. Therefore, we averaged the results of the three questions for each participant. There were no significant differences between the groups (the higher- and lower-power participants) on these measures ($p > .05$). There was a very low amount of emotional conflict according to the measures (averages between 1 and 2 on a scale of 5). Participants indicated moderate task-related conflict in both groups (average 3 out of 5).

On the empowerment and recognition self-report measures, the Cronbach α values were .661 and .371 respectively. Each question was analyzed independently. There was not a significant difference between groups ($p > .05$) on all six questions. The participants reported feeling empowered and that they were able to give and receive recognition (averages between 4 and 5 on a scale of 5).

According to these self-report measures, which were completed following the discussion and final argument, participants disagreed with the confederate over the topic at hand (task-based conflict), but they did not feel angry or emotional tension during the discussion (Jehn's (1995; 2001) conflict scales). Further, they felt as though they could openly express themselves, and they were receptive to the confederate during the discussion (the empowerment and recognition self-report measures). Retrospective self-report, however, can be susceptible to social desirability bias (Heppner et al., 2016). Participants may respond that they were open to the other person, that they did not get angry, etc. because that is the socially desirable answer.

Weakness and alienation are expressed and experienced on a moment-by-moment basis (Bush & Folger, 2010a). The video coding results, which applied static rating scales described above and seen in Appendix B, found times when higher- and lower-participants showed signs of disengaging from the confederate and were unresponsive to certain viewpoints. These scales were rooted in the literature of relationship-focused mediation and conflict theory and looked for prototypical signals as well as explicit indications of disengagement and/or denial of recognition.

There were moderate levels of weakness and alienation during interactions. For an average participant, in both groups, nearly half of the video segments show

weakness and alienation. These results are summarized in Table 5.13 and Figure 5.13. On the video rating measures, there were no significant differences between groups. **The conflict was experienced similarly by lower- and higher-power participants. This is consistent with literature in transformative mediation.**

Further, when participants showed weakness or refused to recognize the confederate's viewpoint, they often maintained the negative state in the next coded video segment. Tables 5.14 and 5.15 summarize the proportion of opportunities for empowerment and recognition shifts in which participants showed signs of empowerment and recognition. The median values, for both groups, indicate that when a participant showed weakness in one video segment, she showed strength in the next video segment less than half the time. When a participant showed a lack of recognition in a video segment, she refused recognition in the following video segment over half the time. **This is consistent with negative conflict described in transformative mediation literature.**

Participants tried to avoid overt conflict or conceal/avoid negative feelings during and after the interaction. None of the participants in either condition mentioned their feelings in the moment (e.g. being upset or off put by the confederate).

Table 5.13: The average participant in both groups had signs of weakness and alienation in nearly half the labeled video segments. This **supports the fact that we were able to drive negative conflict during the discussions and that lower and higher-power dyad members similarly perceived the conflict (in agreement with the literature).**

Proportion of Video Segments Showing Weakness and Alienation				
Proportion of Labeled Video Segments Showing Weakness/Alienation	Two-Sample, Two-Tailed t-test Result			
	Lower-Power Participant Mean (Standard Deviation) (n = 14)	Higher-Power Participant Mean (Standard Deviation) (n = 14)	t-score	p-value
Weakness	0.402 (0.249)	0.492 (0.255)	t(26) = -0.95	p = .351
Alienation	0.496 (0.237)	0.515 (0.259)	t(26) = -0.20	p = .844

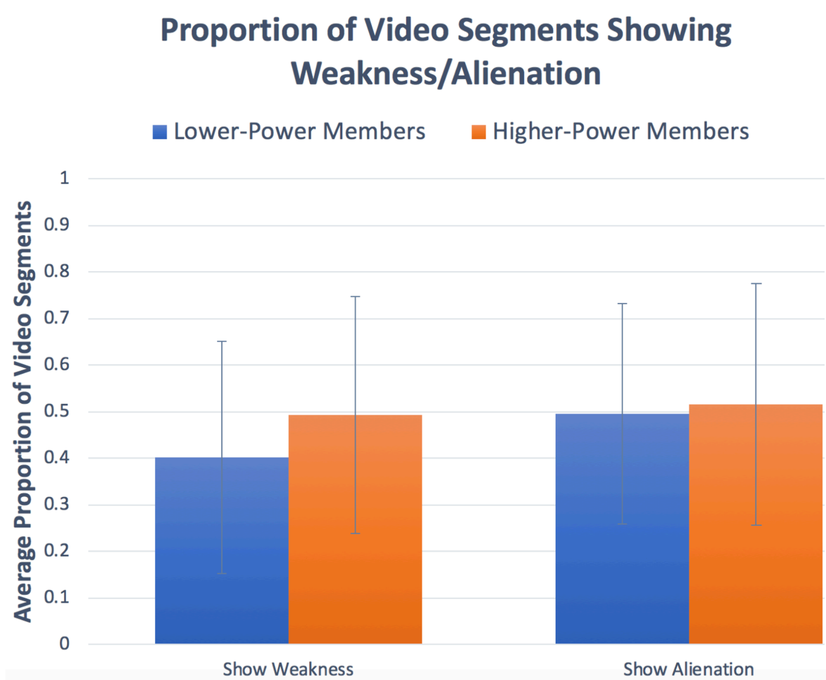


Figure 5.13: The proportion of video segments that show weakness and alienation for both the higher- and lower-power groups. The average participant in both groups had nearly half of the labeled segments showing weakness and/or alienation. **This indicates negative conflict dynamics were present in the discussion and affected higher and lower-power dyad members alike. This fits what is seen in the mediation literature.**

Table 5.14: When a person shows weakness in an interaction, there are subsequent opportunities for that individual to reengage with his/her interaction partner. This shift in behavior is a move of empowerment. This table summarizes the proportion of opportunities for moves of empowerment in which the participants did show signs of reengagement. There was not a significant difference between groups. This agrees with the literature from transformative mediation. **An individual's level of power in a conflict does not change how they experience the cycle of weakness and alienation.**

Measure	Two-Sample, Two-Tailed Mann-Whitney U Test Result			
	Lower-Power Participant Mean/Medium (Standard Deviation) (n = 10)	Higher-Power Participant Mean/Medium (Standard Deviation) (n = 12)	U-Value	p-value
Proportion of Moves of Empowerment to Opportunities for Empowerment	0.521/0.4 (0.399)	0.272/0.167 (0.327)	U = 36.5	p = .123

Table 5.15: When a person refuses recognition in an interaction, there are subsequent opportunities for that individual to show attention and acknowledge the ideas the viewpoint of his/her interaction partner. This shift in behavior is a move of recognition. This table summarizes the proportion of opportunities for moves of recognition in which the participants did show recognition. There was not a significant difference between groups. This agrees with the literature from transformative mediation. **An individual's level of power in a conflict does not change how they experience the cycle of weakness and alienation.**

Measure	Two-Sample, Two-Tailed Mann-Whitney U Test Result			
	Lower-Power Participant Mean/Medium (Standard Deviation) (n = 12)	Higher-Power Participant Mean/Medium (Standard Deviation) (n = 13)	U-Value	p-value
Proportion of Moves of Recognition to Opportunities for Recognition	0.455/0.417 (0.372)	0.377/0.0 (0.462)	U = 62.5	p = .406

The Cronbach α value for the involvement and affection on the Burgoon and Hale (1984; 1987) Relational Communication Scale was 0.8536. Therefore, we averaged the items for each participant for this scale. There was a significant difference between the group with the participant in the lower-power role and the group with the participant in the higher-power role ($p < .05$). Participants in the lower-power position rated the confederate as being more engaged and warmer towards them. This result is summarized in Table 5.16 and Figure 5.14. **This is evidence of the manipulation to create the power dynamic working for the discussion.** The higher-power participants saw the confederate as combative because he was disagreeable when his role was just to help them. The lower-power participants were more willing more appreciative of him listening to any of there ideas because they were there to help.

Table 5.16: Higher-power participants viewed the confederate as colder and more distant than lower-power participants. This indicates the participants internalized their interaction roles. The higher-power participants resented the combativeness of the confederate. The lower-power participants were more appreciative the confederate considered their ideas at all.

Relational Communication Scale Measures (Likert-scales 1 to 7) (Burgoon & Hale, 1984; 1987)	Two-Sample, Two-Tailed t-test Result			
	Lower-Power Participant Mean (Standard Deviation) (n = 14)	Higher-Power Participant Mean (Standard Deviation) (n = 14)	t-score	p-value
Intimacy: Involvement/Affection	5.621 (0.638)	4.813 (0.616)	$t(26) = 3.409$	$p = .002$

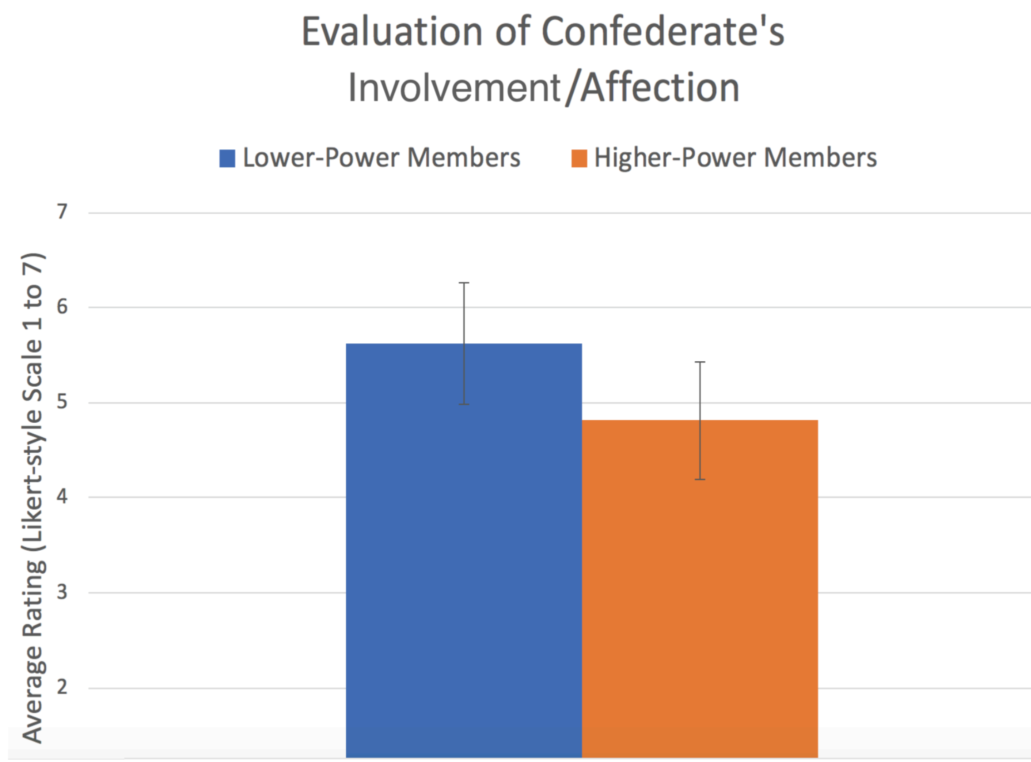


Figure 5.14: The rating of the confederate's perceived affection and involvement in the higher- and lower-power conditions. **The lower-power participants viewed the confederate as more engaged and warmer than the higher-power participants**.

Another Relational Communication Scale measurement assessed the dominance of the confederate (how much control he exerted in the conversation). The Cronbach α value for this scale was 0.0752. The 21 items that composed the scale were analyzed individually using the same process as what has been described. Of the 21 questions, 20 were not significant at a $p = .05$ level ($p > .05$). The only item that was significant at $p = .05$ ($p < .05$) stated directly that the confederate "dominated the conversation". See Table 5.17. Lower-power participants may have seen the confederate as more dominate during the conversation than higher-power participants, which would be an **indication of the participants internalizing the lower- and higher-power roles assigned at the interaction's outset**.

Table 5.17: **Participants seemed to respond to the roles assigned by the experimenter for the interaction**. Lower-power participants rated the higher-power confederate member as more dominate and social during the discussion than the participants who had the higher-power role with the lower-power confederate.

Relational Communication Scale Measures (Likert-scales 1 to 7) (Burgoon & Hale, 1984; 1987)	Two-Sample, Two-Tailed Mann-Whitney U Test Result			
	Lower-Power Participant Median (Standard Deviation) (n = 14)	Higher-Power Participant Median (Standard Deviation) (n = 14)	U	p-value
"Your partner in this interaction dominated the conversation"	4.0 (1.139)	2.0 (0.917)	U = 41.5	p = .010
"Your partner in this interaction was as interested in building a good relationship as completing the task at hand"	6.0 (1.399)	3.5 (1.703)	U = 47.0	p = .020

There were 21 items related to dominance answered by each participant; it is common practice to use correction and lower the level of significance to $p = .0025$ (Miller, 1966) when being conservative. This suggests no significant difference between groups on the dominance scale. It is worth noting, however, the effect size of this item was medium to large (z-score = 2.57307, r =.486). We speculate participants in the lower-power role may have perceived the confederate to take over the conversation more so than those in the higher-power role.

The measure related to the confederate's social versus task-oriented nature had a Cronbach's α value of .133168. The four items were analyzed independently. Three of the four items were not significant at $p = .05$ level ($p > .05$). The item that was significant at $p = .05$ ($p < .05$) stated the confederate "was as interested in building a good relationship as in completing the task at hand". See the second entry in Table 5.3. Again there is the question of using correction, which would make the result not significant at $p = .0125$. Again, the effect size is medium to large (z-score = -2.32, r = .438). We speculate that participants in lower-power roles saw the confederate as more social-oriented than those in the higher-power roles.

The measures related to receptivity/trust, similarity/ inclusion/depth, arousal, and formality from the Relational Communication Scale had Cronbach's α values of .7366, .7682, .8354 and .762 respectively. We averaged the items of these scales for each participant. There was not a significant difference between the two groups on any scale ($p > .05$).

The robot was perceived similarly by both higher- and lower-power dyad members. There were no significant differences between groups ($p > .05$) on the three questions regarding the participants' impressions of the robot. Participants in both groups disagreed that the robot was disruptive (average of approximately 2 out of 5), and the participants in both groups were neutral when deciding if the robot understood how they felt or was on their side (averages of 3 out of 5). There was not a significant difference between groups ($p > .05$). This includes there being no differences in the number of mentions of the robot or glances toward the robot during the discussion.

To summarize, the most critical findings of this section were:

- There were differences between how higher- and lower-power participants perceived the experiment confederate (according to self-report measures). This indicates the experiment manipulation was successful in creating a hierarchy.
- There were not significant differences according to self-report measures in how higher- and lower- power participants experienced the conflict. The weakness and

alienation displayed by higher- and lower-power participants was not significantly different according to video-rating or self-report measures. This is an indication that the robot did not change typical conflict dynamics from developing in hierarchical relationships.

- The robot was perceived and responded to similarly by both higher- and lower-power dyad members. The agent was not biased or disruptive according to self-report and objective measures.

5.7 Study Discussion

This study was in part about gathering baseline data to which data gathered with an intervening robot could be compared, but it served several other purposes as well. It was important to verify that conflict with problematic dynamics could be generated with our study design and that our study was structured in a way that made the power difference between the higher- and lower-power individuals apparent. It also provided information about how the mere presence of a robot effects interactions with destructive conflict and how relationship members perceive such a robot. This relates directly with subsidiary questions one and four from chapter one. The answers to these subsidiary questions are discussed below and summarized in Table 5.18.

The manipulation to set up a power dynamic within the relationship was successful. The higher-power participants saw the confederate as less engaged and colder compared to lower-power members. It makes sense that this is the case. In cases where the confederate was a lower power relationship member, he had no reason to be combative. He was explicitly told to help generate the strongest argument possible, and he immediately began by being disagreeable (dismissing the ideas of the person tasked with actually making the argument). When the participants were lower-power relationship members, they were tasked with helping the confederate. He did not have to listen to their ideas. Their ideas were shown recognition if they showed recognition to the confederate at points later in the discussion. Lower-power participants also rated the confederate as dominating the discussion and more socially oriented than higher-power participants.

The unengaged robot did not stop the negative conflict dynamics from emerging in the hierarchical relationships. The lack of significant differences between groups on the empowerment/recognition measures and Jehn's (1995; 2001) intragroup conflict measures says that the conflict was experienced the same for the higher-power member as the lower-power member. This is consistent with the literature in transformative mediation (Bush & Folger, 2010a; 2010b). The video ratings showed that almost half of the statements by the average participant showed weakness and a lack of recognition. This indicates that ***the confederate was able to generate the type of conflict described in the transformative mediation literature***, and the mere presence of an unengaged robot did not disrupt this cycle of weakness and alienation.

The participants in the higher- and lower-power groups perceived the robot similarly. There were no significant differences between how the lower- and higher-power participant members perceived the unengaged robot according to the self-report measures. The idea that higher- and lower-power relationship members viewed the robot similarly is further supported by objective measures. There was not a significant difference in the amount higher- and lower-power relationship members looked at the robot or commented on the robot.

Qualitative remarks in both groups indicate that they were unsure what the robot's role was in the interaction and ***largely ignored the robot***. There were many participants who commented on the robot fading into the background and the occasional motor noise being the only thing that reminded them that the robot was there.

Table 5.18: Summary of the three subsidiary questions the baseline human-robot study helped to answer.

<table>
<tr><th colspan="2">Answering Subsidiary Questions – Unengaged Robot Study</th></tr>
<tr><td>Subsidiary Question</td><td>Summary of What Was Found</td></tr>
<tr><td>How is an unengaged robot, that has been introduced as something to support communication in the relationship, perceived by the dyad members and how do interventions change this perception?</td><td>• Higher- and lower- power participants had similar perceptions of the robot (based on self-report and objective measures).
  o The robot is something that shrinks into the background, and the dyad members largely ignore it.
    ▪ Participants in both groups did not mention the robot.
    ▪ The average number of glances toward the robot between the groups were not significantly different
    ▪ Participants explicitly commented that they occasionally took notice of the robot, but it was only when its motors made a sound.
  o The participants in both indicated that they were unsure of the role the robot was playing.
    ▪ They average participant was unsure if the robot understood how he/she felt based on a self-report question.
    ▪ Also, the average participant was unsure if the robot was on his/her side during the discussion based on self-report.</td></tr>
<tr><td>What issues does the power dynamic in the relationship present for the robotic agent when trying to provide a conflict process that is viewed as fair and equitable by both participants?</td><td>• The perceptions of the robot were not shaped by the participant’s power in the dyad.
  o As noted above, there was not a significant difference between groups on the attention paid to the robot or the self-reported perceptions of the robot.
• The mere presence of the robot did not change typical negative conflict dynamics from developing when participants had the higher- or lower-power role in the dyad.
  o There was weakness and alienation shown in almost half of the rated video segments for the average higher- and lower- power dyad members.
• It is important for the robot’s interventions to treat the parties the same to avoid making the robot seem biased.</td></tr>
</table>

Table 5.18 (cont.): Summary of the three subsidiary questions the baseline human-robot study helped to answer.

What channels of communication should the agent use when supporting the dyad, how overt should these communications be in order to avoid upsetting the relationship, and how should the agent choose between interventions?	• The unengaged robot simply faded into the background for many participants.
	  o Participants rarely glanced at the robot (averages in both groups were fewer than once per minute)
	  o Participants explicitly stated that the robot was forgotten. They only took notice of the robot when the noticed a loud motor noise.
	• Interventions need to be overt enough to attract attention or at least glances from the dyad without continuously disrupting the flow of the conversation.

5.8 Identifying States of Weakness and Alienation

The second subsidiary question from chapter one asks how a hierarchical relationship should be represented to understand when there are problematic dynamics and support positive change. In chapter four, we gave a partial implementation of the computational model that allowed a robot to identify two problematic states. The states were the negativity state and insensitivity state.

Our autonomous system identifies negativity (a state of weakness) and insensitivity (a state of alienation) using some of the same cues as professional mediators when identifying strain in human-human relationships (Bush & Folger, 2010a; Bush & Folger, 2010b; Moen et al., 2001; Retzinger, 1991; Tickle-Degnen, 2006). There were four percepts identified in chapter four: the degree to which the thinking (speech) of the participant was fragmented, the tempo of her speech, the loudness of her speech, and the participant's orientation toward or away from the confederate. The thresholds used to set these percepts were learned from the data collected in this study. A summary of how the thresholds for the percepts were learned is given in Table 5.19.

Table 5.19: The process by which the experimenter set the thresholds for the percept generating functions described in detail in section 4.2.

Procedure to Train Computational Model	
Threshold Setting Steps	Description
Step 1: Label Video Segments	• The experimenter and video coder rated segments of video for the states of negativity and insensitivity. • They indicated why a state was active or inactive in a given segment (following the guidelines given in Appendix B). • For example, they may have decided the negativity state was active in particular video segment because the participant was speaking quickly and oriented away from the experiment confederate for a long period of time.
Step 2: Gather Processed Sensor Values From Labeled Video Segments	• The experimenter ran the code (fully described in section 4.2) that processes the raw sensor values (from the microphone and camera) on the labeled video segments. • This code outputs: loudness values, the number of syllables said per second in each utterance, the number of utterances said per minute, the length of each utterance, and the number of frames in which the participant was oriented away from the confederate. • Speaking loudly/quietly, speaking fast, speaking in short utterances, being oriented away for extended periods are indicative of negativity and insensitivity being active (Bush & Folger, 2010a; Retzinger, 1991). • The experimenter was able to see the loudness, pacing, and orientation values for sections of video labeled as problematic (i.e. the experimenter knew the participant was speaking loudly, speaking fast, etc.) in these video sections. • These values were ground truth values for what is loud, fast, etc.
Step 3: Examine Sensor Values	• The processed sensor values for the video segments were examined within each individual to ensure that the pacing, loudness, orientation values were different for sections of video where the coders indicated problematic communication styles. • Median values were compared between participants for sections of video where pacing was fast, voice was loud, etc. • Segments with pacing, orientation and fragmentation issues showed similar values between participants. • Loudness values were different between participants.
Step 4: Set Threshold Values	• The thresholds for pacing, orientation and fragmentation were set as median values. • The thresholds for loudness depend on individual baselines.

Just as for the video rating measures above, one experimenter and one video coder, who was unaware of the study's purpose and hypotheses, rated the videos using guides rooted in transformative mediation literature (Bush & Folger, 2010a; Bush & Folger, 2010b; Moen et al., 2001; Retzinger, 1991; Tickle-Degnen, 2006). They rated segments of video for the negativity and insensitivity states. The guide used by the coders appears in Appendix B.

The video coders gave a binary rating for the negativity state (of active or inactive) for each speaking part of the lower-power participants (using the guidelines provided in Appendix B). They gave a binary rating for the insensitivity state (of active or inactive) for the speaking and listening parts in higher-power participants (also using Appendix B). As above, Cohen's κ was used to assess inter-rater reliability. The κ value for the negativity ratings was .526. The κ value for insensitivity was .551. This signifies moderate agreement between raters (Simand & Wright, 2005) and is similar to reliability estimates in previous studies (e.g. Jung, 2016).

When a state was labeled as active, the video coders indicated why that state was active. They indicated that the person was speaking loudly, rapidly, using language that alienated the other, etc. As described in Table 5.18, this data served as a ground truth to set the threshold values in the percept-generating functions (Algorithms 4.1-4.4). The median values relating to speech pacing, orientation, and fragmented speech in problematic video sections were used to classify the participant's speech as fast or fragmented and the orientation as withdrawn or intense. Problematic video sections here is referring to when there are relationship problems. These are defined as the sections of video in which the states of weakness (negativity) or alienation (insensitivity) were present in the relationship by the video coders. The ways in which these values are derived are given in section 4.2. There were 102 examples of negative communication patterns within the 28 participants run as part of this study. This is similar to the number of trials run in previous related studies (e.g. Hoffman et al., 2015; Jung, 2016). The thresholds appear in Table 5.20.

Table 5.20: The threshold values for the percept-generating functions (Algorithms 4.1-4.4).

Thresholds Used to Set Percept Values in Current Model Implmentation		
Percept	Description	Threshold Value
Pacing	Fast Tempo	> 4.3 syllables per second
	Raised Tempo	> 3.6 syllables per second
	Calm Speech	<= 3.6 syllables per second
Loudness	Silent	No openSMILE loudness value exceeds 0.1 for 30 seconds.
	Quiet Speech	The average openSMILE loudness value for the last 30 seconds of speech is more than 1 standard deviation below the individual's baseline loudness (the baseline is set as described in chapter 4.2 using the individual's first 20 seconds of speech).
	Loud Speech	There were average openSMILE loudness value for the last 30 seconds of speech is more than 1.25 standard deviations above the individual's baseline loudness
	Stressed Speech	The individual makes multiple sounds of at least one-half second over the course of the most recent 30 seconds that exceed 2 times the baseline loudness or 5 sounds that exceed 1.25 times the individual's baseline loudness.
	Calm Speech	The speech does not fit into the other categories (within the range defined by the baseline).
Fragmented Speech	Not Fragmented Speech	Average utterance is above 2.25 seconds or Average utterance is above 1 second and the utterance per minute are below 7
	Slightly Fragmented Speech	The average utterance is above .8 seconds and the utterances per minute is below 25
	Highly Fragmented Speech	The other conditions are not met for the speech being not fragmented or slightly fragmented. There are many utterances per minute, and they are very short.

Table 5.20 (cont.): The threshold values for the percept-generating functions (Algorithms 4.1-4.4).

Aversive Orientation	Glaring/Frustrated	Oriented and leaning toward the confederate for more than 80% of the previous 45 seconds Or Oriented toward the confederate for more than 95% of the previous 45 seconds.
	Shameful	Oriented away from the confederate for more than 80% of the previous 45 seconds.
	Normal	Does not fit into the previous two categories.

The experimenter played the audio-visual recordings back in real time and had the computational model (described fully in chapter 3 and section 4.2) label the weakness and alienation states for each participant to evaluate its performance on the training data set. This algorithm labeled negativity as active or inactive during speaking parts in trials where the participant was the lower-power relationship member. It labeled the insensitivity state as active or inactive during each speaking and listening part when the participant was the higher-power dyad member. The algorithm's label was active if it the state was set active at any point during the section of video and inactive if the state was never set active during the section of video. This algorithm is fully described in chapter 3 and section 4.2 (Algorithms 4.1-4.4, Table 4.9). Section 3.3 provides a step-by-step example of how the computational model sets the states active or inactive. As noted in chapter four, the entire code base is available online[15].

The video coders' ratings served as a ground truth for this algorithm. If both video coders agreed that a problematic relationship state was active or inactive in a particular video segment, then the rating given by the algorithm was compared to the rating given by the video coders for that segment. If the video coders did not agree on the rating for a certain video segment, then this video segment was discarded from the analysis. The results of this process are summarized in Table 5.21.

Previous research that has labelled negative emotional displays using multiple video coders has sometimes chosen one of the coders results at random to include in the analysis (e.g. Jung, 2016) rather than discarding segments of disagreement. Because one of the video coders in this study was an experimenter, it was important to limit bias and ensure adherence to the rating scales given in Appendix B. Appendix E provides the steps that were taken to ensure this adherence.

---

[15] https://www.cc.gatech.edu/ai/robot-lab/nri_thesis/code/All_Code.zip

Table 5.21: The precision, recall and specificity for the algorithm on the training data. There is a high true negative rate. The true positive rate, however, is low. **The algorithm has moderate positive predictive value using the nonverbal and paralinguistic behaviors that it currently incorporates.**

Outcome Measures for Current Model Implementation on Training Set			
State	Precision	Recall	Specificity
Negativity	.42	.385	.714
Insensitivity	.516	.211	.75

The experimenter also examined the proportion of video segments that had negativity/insensitivity labeled as active by the experimenter and the video coder and the proportion of video segments that had negativity/insensitivity labeled as active by the algorithm for each participant. As noted above, the algorithm labeled the insensitivity state as active or inactive during each speaking and listening part of the discussion when the participant was the higher-power dyad member. It labeled negativity as active or inactive during each speaking part of the discussion when the participant was the lower-power dyad member. If the negativity/insensitivity was set active at all during a particular video segment, then the algorithm labeled that segment active. If the algorithm never set the states active, then the algorithm labeled the video segment inactive. The video coders rated all of the same video segments as active or inactive.

This was to see if the algorithm's ratings correlated with the video coders' ratings during the discussions. A correlation between the algorithm and the video coders would indicate that the number of time that these nonverbal and paralinguistic cues appeared is related to the number of expressions of weakness and alienation by participants during the discussion. The results for this analysis are summarized in Table 5.22 and 5.23 and Figures 5.15 and 5.16. In Table 5.22, the proportion of video segments in which the negativity state was labeled as active by the algorithm and by the video coders is shown for each participant. In Table 5.23, the proportion of video segments in which insensitivity was labelled as active by the algorithm and the video coders is shown for each participant.

There were not significant correlations between the labels of the algorithm and the labels of the video coders for negativity (Spearman's $\rho(12) = 0.319$, two-tailed p-value = 0.266) or insensitivity (Spearman's $\rho(10) = 0.004$, two-tailed p-value = 0.991). **The lack of correlation says that the features that we have identified as relating to weakness and alienation within the relationship are not necessarily going to indicative of how often participants are displaying weakness and alienation. They can be weak and consistently trying separated from the other without displaying these signals often. Further, the cues, when displayed, only have moderate predictive power. Sometimes a larger context is needed to understand these cues (or the absence of these cues).**

Transformative mediators respond in moments where they are able to highlight a particular point that a dyad member can clarify, a point that is important to a dyad member, or a decision point (Bush & Folger, 2010b; Jorgensen et al., 2001). In these moments, the identification of these states is highly dependent on verbal cues as opposed to the nonverbal cues that our agent could identify. As natural language processing technology matures, artificial intelligence will be capable of identifying these instances.

Further, the paralinguistic and nonverbal cues that the autonomous system is using to identify these states may appear but not persist in an interaction and may not be present during utterances that show separation in the dyad (Retzinger, 1991). These nonverbal cues are indicators of the states but are not synonymous with the states.

Technical issues also affected the performance of the algorithm. For example, participants (especially higher-power participants) would take notes when listening to the confederate's responses. The participants would look almost straight down in order to take the notes, totally obscuring the view of their faces for large amounts of time. As described in chapter four, the algorithm set the problematic relationship states active when there were multiple cues that were indicative of the state being active. When the face was not visible, the autonomous system cannot use the orientation cue to set the problematic states as active or inactive. Therefore, it only depended on the use of the other (paralinguistic) cues.

As technology matures, additional cues, such as cues taken from natural language, can be added to make the system more robust to missing information. Technical failures can limit the recall (the number of relevant elements identified) of the system. The robotic system is looking for cues that indicate active states of weakness (negativity) or alienation (insensitivity). A technical issue can prevent the robot from identifying these cues, which lessens its ability to identify the states as active.

Table 5.22: The proportion of the video segments in which negativity was labeled active for each lower-power participant by the video coders and the autonomous system. It shows a lack of correlation between the autonomous system and video coders. The nonverbal and paralinguistic behaviors used by the autonomous system are indicative of the presence of negativity in the relationship (e.g. Moen et al., 2001; Retzinger, 1991). The number of video segments in which the participant withdrew or pushed away the experiment confederate (those labeled as active by the video coders), however, were not correlated with the video segments in which the participant displayed the nonverbal and paralinguistic behaviors associated with the withdrawal and/or frustration (those labeled as active by the algorithm). **There are participants who do not behave in this way when separating from the confederate. There are going to be participants who demonstrate these behaviors but remain engaged with the confederate. These cues only have moderate positive predictive power.**

Participant	**Proportion of Video Sections with Negativity Active**	
	Video Coders	Autonomous System
1	.875	.5
2	.2	0
3	0	0
4	.143	.143
5	.3	.2
6	.444	.222
7	.571	.143
8	.389	.222
9	.6	.4
10	.083	.429
11	.167	.833
12	.778	.667
13	.25	.5
14	.389	.778

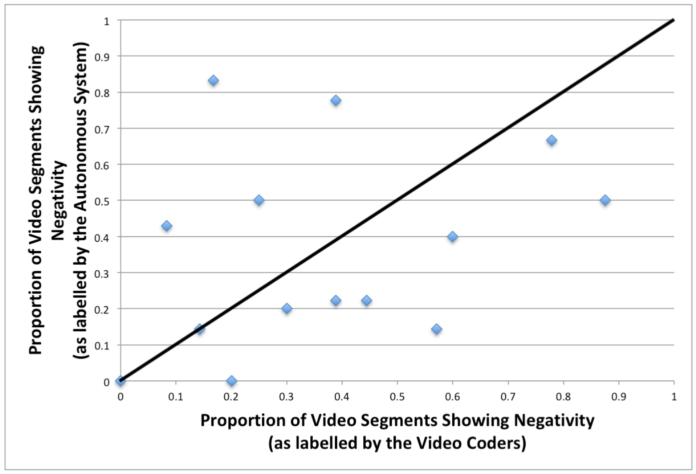


Figure 5.15: The amount of video segments in which a participant pushed away or withdrew from the confederate (the segments labeled as having active negativity by the video coders) were not correlated with the amount of video segments in which the participant displayed nonverbal and paralinguistic behaviors associated with negativity (those labeled as having active negativity by the algorithm). This lack of correlation implies that although these behaviors are associated with the presence of negativity in the relationship (Moen et al., 2001; Retzinger, 1991), these behaviors, however, do not appear at the same rate participants displayed negativity. **Certain participants will not display these behaviors and separate from the other. Further, these cues on their own only have moderate positive predictive power.**

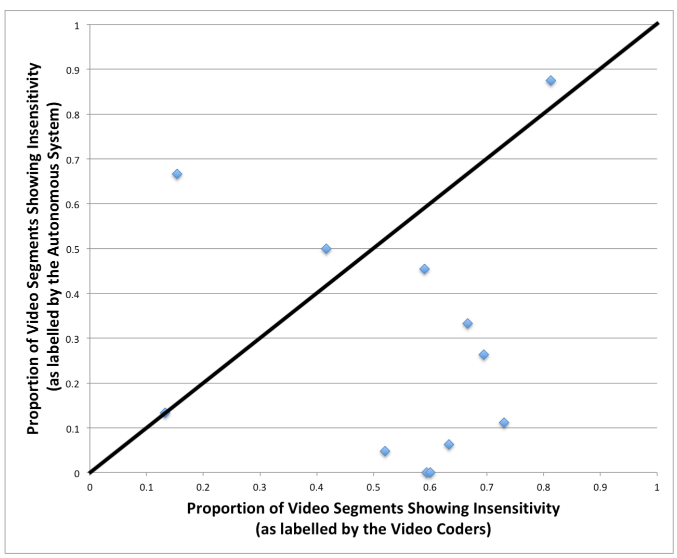


Figure 5.16: The proportion of video segments labeled as showing insensitivity by the algorithm was not significantly correlated with the proportion of video segments labeled as active by the video coders. The nonverbal and paralinguistic cues used by the algorithm identified the presence of insensitivity within certain participants but not all participants. **Certain participants displayed these behaviors while having empathy for the confederate, while others did not display these behaviors but were alienated.**

Table 5.23: The proportion of the video segments in which insensitivity was labeled active for each higher-power participant by the video coders and the autonomous system. It shows a lack of correlation between the autonomous system and video coders. The nonverbal and paralinguistic behaviors used by the autonomous system are indicative of the presence of insensitivity in the relationship (e.g. Moen et al., 2001; Retzinger, 1991). The number of video segments in which the participant refused to recognize the viewpoints of or was disengaged from the experiment confederate (those labeled as active by the video coders), however, were not correlated with the video segments in which the participant displayed the nonverbal and paralinguistic behaviors associated with hostility or indifference (those labeled as active by the autonomous system). **There are going to be participants who do not show these behaviors even when they lack empathy for the other in the moment. Further, these cues on their own only have moderate positive predictive power.**

Participants	**Proportion of Video Sections with Insensitivity Active**	
	Video Coders	Autonomous System
1	.417	.5
2	.694	.263
3	.133	.133
4	.8125	.875
5	.666	.333
6	.52	.0476
7	.594	0
8	.633	.0625
9	.59	.4545
10	.73	.111
11	.154	.666
12	.6	0

The difficulty of identifying these states in the moment is discussed at length in the transformative mediation literature, and there are guidelines that have been proposed for new practitioners of transformative mediation to support relationships even when they are unsure of if there is acute weakness and alienation being expressed (Jorgensen et al., 2001). These are the intervention guidelines that we adopted for the intervening robot that was tested. Table 5.24 summarizes how the analysis relates to the subsidiary questions.

The algorithm analyzed as part of this chapter was incorporated into an autonomous robotic system. This robot has been tested in a study described in the following chapter. The algorithm is using certain cues that are used by transformative mediators to identify when a relationship is strained (Bush & Folger, 2010a; Bush & Folger, 2010b; Moen et al., 2001; Retzinger, 1991; Tickle-Degnen, 2006). The robot intervenes in ways that follow the guidelines provided by the transformative mediation literature (Bush & Folger, 2010b; Jorgensen et al., 2001). These results will be compared to the results of this study to see if such a robot is able to support positive shifts in the hierarchical human-human relationships.

Table 5.24: Summary of how the work done to train the autonomous system relates to the subsidiary questions asked in chapter one.

Subsidiary Question	What Study Revealed in Relation to the Subsidiary Question
How can a social robotic agent represent a dyad's relationship state, problematic or otherwise, and decide when to intervene in the relationship?	• The percepts our autonomous system identifies (speech that is of irregular tempo and volume, fragmented speech, and aversive/aggressive orientation) are sufficient for identifying our negativity and insensitivity states in certain relationships, but they are insufficient for identifying negativity and insensitivity in all relationships. o There are relationships in which the algorithm labeled the states appropriately, but there was not a significantly correlation between the video coders' labels and the algorithms labels with the presence of the states across discussions. o The computational model that uses these behavioral cues only has moderate positive predictive power. • Human communication is complex and singular cues are often uninformative or given too much importance (Bush & Folger, 2010a; Bush, 2010; Noce, 2010). o This is something that humans often struggle with and why even professional human mediators may miss certain opportunities for intervention or misread the state of certain relationships (Moen et al., 2001). o Artificial agents will need to continue to incorporate additional signals as technologies mature to accurately identify these states.

Table 5.24 (cont.): Summary of how the work done to train the autonomous system relates to the subsidiary questions asked in chapter one.

What channels of communication should the agent use when supporting the dyad, how overt should these communications be in order to avoid upsetting the relationship, and how should the agent choose between interventions?	• Interventions must be nonjudgmental and inoffensive
	  ○ The robot should begin very subtle in its interventions to avoid unnecessary relationship disruption in the case it is incorrect.
	    ▪ Even professional mediators, especially those early in their career, struggle to identify moments in which weakness and alienation are being expressed (Jorgensen et al., 2001).
	• Human mediators can be successful at supporting the dyadic relationships with subtle and simple interventions (Jorgensen et al., 2001).

This section analyzed the computational model used to identify the problematic relationship states of negativity and insensitivity in the participant-confederate discussion. As described in section 4.2, this model uses nonverbal and paralinguistic behavioral cues that are commonly used by relationship-focused mediators and relationship experts trying to understand problematic communication dynamics (Bush & Folger, 2010a; Bush & Folger, 2010b; Moen et al., 2001; Retzinger, 1991; Tickle-Degnen, 2006). When testing on the data gathered as part of this study, the model has only moderate positive predictive value. It has a high true negative rate, but it has a low true positive rate. As technology matures, additional cues (such as natural language processing) will need to be incorporated to improve the sensitivity of this model.

An autonomous agent that uses the current version of the model may still be able to support communication in hierarchical human-human relationships. Jorgensen et al. (2001) discuss how mediators early in their career can have trouble identifying weakness and alienation in human-human relationships. These mediators use subtle supportive interventions to help the human-human relationships. The robotic interventions implemented and tested in the following chapter use the guidelines provided by Jorgensen et al. (2001) for early-career mediators.

This analysis has highlighted three important points for our autonomous robotic system:

- The nonverbal and paralinguistic cues that the system incorporates have moderate positive predictive value. That is, when an individual displays these cues in a certain moment, there is a moderate likelihood he/she is separating from or already alienated from the other individual.
- These cues have been shown to be indicative of weakness and alienation within an individual (e.g. Moen et al., 2001; Retzinger, 1991). The lack of the correlation between the artificially intelligent system and video coders, however, says that the amount an individual displays these cues are not informative about the amount he/she tries to separate from the other or the amount he/she demonstrates a lack of empathy for the other.
- Due to the limitations of these cues, and the fact that the technologies that can improve these results are still maturing, our autonomous system's interventions are informed by guidelines designed for new relationship-focused mediators who also have trouble identifying states of weakness and alienation (Jorgensen et al., 2001).

5.9 Conclusions

As summarized in Tables 5.17 and 5.22, this chapter presented a baseline human robot interaction study that helps to provide insight into all four of the subsidiary questions asked in chapter one. This baseline study saw a discussion take place between two people where there was a hierarchy established and negative conflict took place. In this scenario, an unengaged robot was present that was introduced as something that could help with the dyads communication. The robot made subtle movements that were not correlated with the conversation in any way.

The next chapter in this thesis presents the findings of a follow-up study with an intervening robotic agent. It follows the same procedure as the study presented in this chapter, but the robotic agent that was present in this study now supports the relationship using the computational model and the interventions introduced in chapter four. The robot identifies states of negativity and insensitivity in the confederate-participant relationships using the cues that have previously been introduced and which are rooted in transformative mediation literature (Bush & Folger, 2010a; Bush & Folger, 2010b; Moen et al., 2001; Retzinger, 1991; Tickle-Degnen, 2006).

The robot then intervened using interventions that are based in the tenants of transformative mediation (Bush & Folger, 2010b; Jorgensen et al., 2001). This second study provides results that can be compared to the study presented in this chapter to see if an intervening robotic agent can support positive shifts (shifts of empowerment and recognition) in hierarchical human-human relationships.

CHAPTER 6

INTERVENING AUTONOMOUS ROBOT STUDY

When two people in a hierarchical relationship are engaged in negative conflict, the mere presence of a robot does not disrupt typical conflict dynamics from developing. The previous chapter presented a study with an unengaged robot at the periphery of an interaction between an experiment confederate and participant. The robot moved subtly and in ways that were not correlated with the conversation. Both lower- and higher-power participants showed moderate levels of weakness and alienation during these discussions. This chapter presents results from an experiment with an intervening robot that intervened to help ameliorate the problematic conflict dynamics and the dyad's communication.

The chapter begins by highlighting the difference between this study and the previous study. Subsequently, the results of a pilot study are presented; this pilot study helped to shape the robot's interventions. The third section summarizes the results from the demographics and opinion surveys given to participants at the outset of the experiment. The chapter concludes by presenting the results of the study on each of the study measures and discussing how the study has helped to answer each of the four subsidiary questions asked in chapter one.

6.1 Intervening Autonomous Robot Study Overview

The procedure in this study is identical to the procedure of the study in chapter five (shown in Table 5.11). The only difference between this study and the previous study is the robot's behavior. The robot remained unengaged in the first study. It moved subtly and randomly in ways not correlated with the dyad's conversation. In this study, the robot tried to identify strained relationship states (as detailed in section 4.2) and intervened (as described in section 4.3). It maintained the unengaged behavior when it was not intervening to ameliorate the relationship. The unengaged robot's finite-state automaton (FSA) is shown in Figure 6.1. The intervening robot's FSA is shown in Figure 6.2.

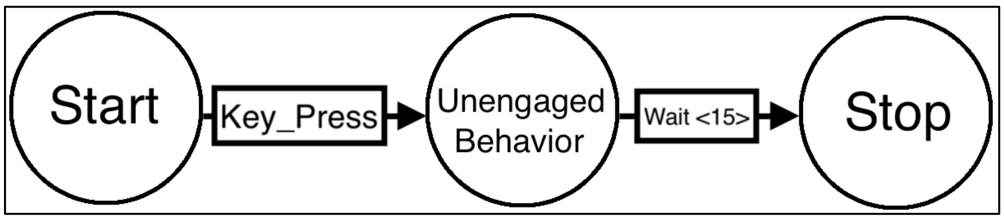


Figure 6.1: FSA for the unengaged robot described in chapter 5. As soon as the experimenter pressed a key, the robot moved subtly and in ways unrelated to the conversation. The experimenter pressed this key when he was introducing the robot to the dyad. The robot maintained this behavior for 15 minutes (until the end of the conversation between the dyad).

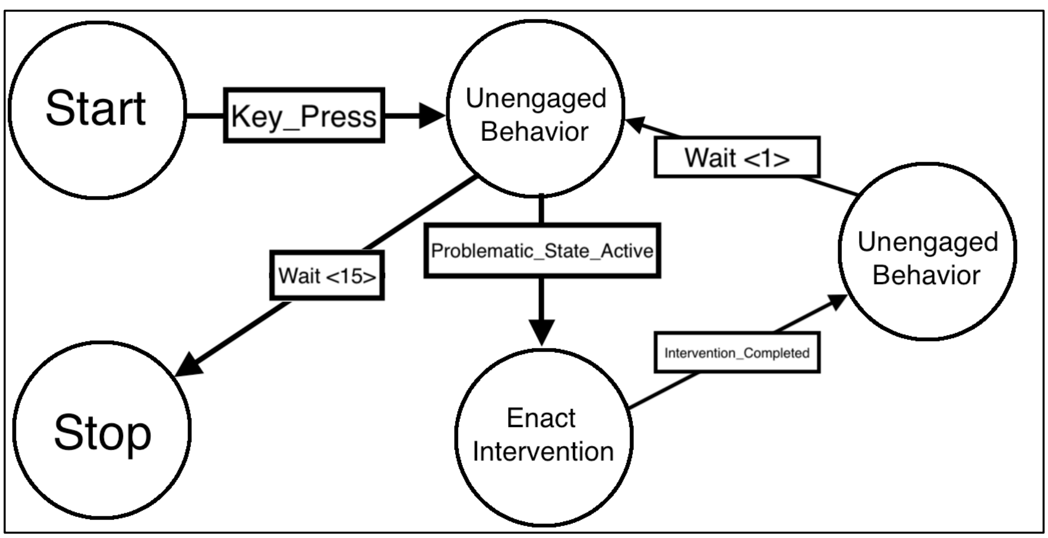


Figure 6.2: FSA for the intervening robot. The robot begins the unengaged behavior when the experimenter presses a key, just as in the unengaged robot study. The robot maintains this behavior until a problematic relationship state becomes active (as described in section 4.2). As soon as the state is active, the robot will enact an intervention (described in section 4.3). When the intervention is completed, it will maintain the unengaged behavior for at least a minute to avoid intervening continuously in the relationship. It will stop at the end of the dyad's conversation (after 15 minutes).

6.2 A Pilot Study to Test Robotic Interventions

As noted in chapter four, the first five participants of this study were set aside to be part of a pilot. We wanted to ensure that the autonomous agent's interventions were straightforward, subtle (to avoid disruption) and did not make participants perceive the agent as biased. As detailed in chapter four, the autonomous robot had to uplift the participants when the negativity state was active; it had to encourage empathy when the insensitivity state was active.

The negativity state was active in the relationship when a lower-power participant tried to separate herself from the higher-power confederate (by withdrawing or pushing the other away in frustration). The insensitivity state was active when a higher-power participant was inattentive or demonstrated hostility toward the lower-power confederate. Tables 6.1 and 6.2 summarize the interventions used as part of the pilot and the transformative mediation guidelines in which they were rooted.

These interventions were used for the first three pilot participants. All three of the participants indicated that they were only aware of the verbal intervention (as shown in Table 6.3). For the final two pilot participants, a short verbal utterance (akin to "ahem") was inserted before the nonverbal interventions. These final two participants were aware of the nonverbal interventions once this short verbal utterance was inserted. All five pilot participants' impressions of the robot are summarized in Tables 6.3 and 6.4.

The design guidelines learned from the five pilot study participants are presented in Table 6.5. The experimenters updated the final interventions based on the design guidelines learned from the pilot. The final interventions are summarized in Tables 6.6 and 6.7 along with justification for why that intervention supports the relationship.

Table 6.1: These were the interventions used to uplift the lower-power participants when the negativity state was active in the relationship. These interventions exactly were used for the first two lower-power pilot participants. A short utterance akin to "ahem" was inserted before strictly nonverbal cues for the final lower-power participant. These are based on the guidelines provided by Jorgensen et al. (2001) for new practitioners of transformative mediation.

Interventions Tested During the Pilot Study to Uplift Lower-Power Dyad Member	
Robotic Intervention	Transformative Mediation Guidelines
$\beta_{uplift,subtle}$ {<Nonverbal cues> [Nonverbal Cue Order] [N1] Attend to the speaking party. [N2] Periodically nod at party to indicate attention.}	Focused attention and nods (or other minimal encouragers) can keep individuals talking such that they express themselves fully. A participant can open up about how she is feeling in the interaction and more fully express her ideas, which is empowering.
$\beta_{uplift,average}$ {<Nonverbal cues> [Nonverbal Cue Order] [N1] Attend to the speaking party. [N2] Mirror the affect of the speaker to show understanding}	When an individual knows she is understood, it can be empowering because she knows that she is conveying herself accurately. Also, her thoughts and feelings are being considered as part of the interaction.
$\beta_{uplift,strong}$ {<Nonverbal cues> [Nonverbal Cue Order] [N1] Attend to the speaking party (weak participant). [N2] Periodically nod at party to indicate attention. [N3] Wait until the participant has stopped speaking before using one verbal cue. When using the verbal cue, maintain gaze toward weak party. Verbal cues give weakened party chance to state self clearly and decision-making power, which are empowering. <Verbal cues> [Verbal Cue Options] [V1] "Based my computations, you seem upset. If this is correct, is there something that can change in this interaction to make you feel better?" [V2] "Before you both get back to your exchange, I'm curious as to how you would use the discussion thus far in the final argument?"}	This is a more overt indicator of understanding. It amplifies the emotion in the room, so that it is easier for either party to address it. Understanding on its own is empowering, but it is also empowering for a person to have the opportunity to make a clear statement about how she is feeling in the moment (to give the person who is present the opportunity to address them). If there are negative emotions, it can be empowering to make a clear decision about what would ameliorate those emotions.

Table 6.2: These were the interventions used to encourage empathy in the higher-power participants when the insensitivity relationship state was active. For the first higher-power pilot participant, these were the exact interventions. For the second higher-power pilot participant, a short utterance akin to "ahem" was added before strictly nonverbal interventions. These are based on the guidelines provided by Jorgensen et al. (2001) for new practitioners of transformative mediation.

Interventions Tested During the Pilot Study to Encourage Empathy in the Higher-Power Dyad Member	
Robotic Interventions	Transformative Mediation Guidelines
$\beta_{empathy,subtle}$ {<Nonverbal cues> [Nonverbal Cue Order] [N1] Attend to the speaking party. [N2] Periodically glance at other party to "check in".}	It is critical for a mediator to frame the mediation session as a discussion between the two parties. Glancing periodically at the other individual reminds the speaker to consider who she is speaking to and reminds the speaker to share the discussion with that other person.
$\beta_{empathy,average}$ {<Nonverbal cues> [Nonverbal Cue Order] [N1] Attend to the speaking party. [N2] Periodically glance at other party to "check in". [N3] During "check in" mirror the affect of the speaker to amplify conversation and allow for recognition.}	Amplifying the affect in the room provides the other party with another opportunity to see how the other individual is feeling and more opportunity to address it. When the negative emotion in the room is addressed directly, and each dyad member can see that the other is committed to working through the negative emotion, it is empowering to both individuals.
$\beta_{empathy,strong}$ {<Nonverbal cues> [Nonverbal Cue Order] [N1] Attend to speaking party (confederate that is being shown a lack of empathy). [N2] Wait until the party has finished speaking before verbal cue. The verbal cue is directed at member that is being shown a lack of empathy. <Verbal cues> [Verbal Cue Options] [V1] "Based my computations, you are very emotional about the current discussion point."}	This allows the dyad member showing the lack of empathy to listen to the other person in the discussion from a distance and hear how she is feeling. It also gives the individual experiencing the negative affect to make a direct and empowered statement about how she is feeling. If a mediator tries to force an individual to see how someone else is feeling, it removes that person's agency and weakens her. If the people having the discussion can make their own decisions about how they express themselves and how they respond to one another, they are granted agency and empowered.

Table 6.3: The pilot study participants' comments on the robot (written or verbally reported to the experimenter after the discussion).

Participants' Comments Regarding the Interventions		
Participant	Comments	Participant Order
Lower-Power 1	Participant: "Robot only spoke once and seemed to have no idea what we were saying to each other."	1
Lower-Power 2	Told experimenter that he/she only noted one intervention.	3
Lower-Power 3	Participant: "I did not know why it did some of the things it did e.g. make a face or raise its hand." Commented to the experimenter after the fact that he thought it was saying he was sad, but he wasn't.	5
Higher-Power 1	Participant: "It only intervened once when my companion was speaking passionately."	2
Higher-Power 2	Noted the "ahem" before the nonverbal interventions.	4

Table 6.4: The pilot study participants' impressions of the robot on Likert-style scale measures.

Participants' Impressions of the Robot				
Participant	The robot was disruptive.	The robot clearly didn't understand how I felt.	The robot was on my side.	Participant Order
Lower-Power 1	Unsure	Unsure	Unsure	1
Lower-Power 2	Disagree	Agree	Agree	3
Lower-Power 3	Strongly Disagree	Agree	Unsure	5
Higher-Power 1	Strongly Disagree	Strongly Disagree	Unsure	2
Higher-Power 2	Unsure	Agree	Unsure	4

Table 6.5: Design guidelines learned from pilot participants.

Design Guideline	Evidence for Design Decision
It is important to simplify verbal utterances made by the robot as much as possible.	The lower-power participants stopped speaking and tried to listen to the robot during its verbal interventions, but it was clear from reviewing the video that they had trouble understanding what the robot said. None of these participants responded to the robot.
The robot should amplify the emotion of the discussion without addressing the individuals directly.	All of the lower-power participants indicated that the robot did not know how they felt (on the Likert-scale measure or in the written comments, Tables 6.3 and 6.4). The robot used the verbal intervention in all three trials. The video coders agreed that there was weakness and negative affect in these trials. The participants would not acknowledge that negative affect during or after the discussion. The first higher-power individual believed that the robot understood the emotion during the interaction. The robot pointed out the (present) negative emotion by addressing the experiment confederate. Making an individual overly self-aware in a public setting can cause that individual to limit personal expression (Joinson, 2001). The robot can amplify what is present in the discussion without directing it at an individual. This also allows the robot to treat the dyad members the same (Bush & Folger, 2010a).
Verbal utterances can be used to attract attention to the robot before a nonverbal intervention.	The first three participants indicated that they had not taken any notice of the strictly nonverbal interventions (Table 6.3). The final two participants did take note of nonverbal interventions when they were preceded with an "ahem" type verbal utterance. It is important for individuals to take note of the interventions. To influence behavior subconsciously is supplanting the decision-making power of the individual; this goes against the fundamental guiding principle in transformative mediation (e.g. Bush & Folger, 2010b)

Table 6.5 (cont.): Design guidelines learned from pilot participants.

Design Guideline	Evidence for Design Decision
Verbal utterances can be used to clarify nonverbal interventions.	One of the participants who took note of a nonverbal display was confused as to what was being communicated (Table 6.3). If a nonverbal display occurs simultaneous with a verbal statement, participants may associate nonverbal displays with that statement in the future.

Table 6.6: The final uplift interventions that were used in the study. These were updated based on the pilot study as well as the guidelines given in the transformative mediation literature, particularly those given by Jorgensen et al. (2001) for new practitioners of transformative mediation. Based on the pilot study, short verbal utterances were inserted before nonverbal interventions to attract participant attention. Verbal interventions were introduced earlier to help frame nonverbal interventions.

Final Uplift Intervention Descriptions and Rationale	
Robotic Interventions	Rationale for Intervention
$\beta_{uplift,1}$ {<Nonverbal cues> [Nonverbal Cue Order] [N1] Attend to the speaking party. [N2] Smile at speaking party. [N3] Periodically nod at party. [N4] Attend to the other party when speaker changes. [N5] Smile at new speaker. [N6] Periodically nod at party to indicate	The robot treats each individual the same. Smiling and nodding are used as encouragement for the parties to express themselves fully.
$\beta_{uplift,2}$ {<Verbal cues> [Verbal Cue Order] [V1] "Ahem" <Nonverbal cues> [Nonverbal Cue Order] [N1] Attend to the speaking party. [N2] Smile at speaking party. [N3] Periodically nod at party. [N4] Attend to the other party when speaker changes. [N5] Smile at new speaker. [N6] Periodically nod at party.}	There is a verbal "ahem" added for the robot to draw attention to itself, so the parties are alerted to the attentive and now responsive agent.
$\beta_{uplift,3}$ {<Nonverbal cues> [Nonverbal Cue Order] [N1] Wait until there is a break in the discussion. [N2] Smile <Verbal cues> [Verbal Cue Order] [V1] "I'm really enjoying hearing both of your insightful ideas"}	The robot is explicit about trusting the dyad in their competence and ability to handle the conversation. When a mediator shows she trusts the parties' abilities to make decisions, it can be empowering for dyad.

Table 6.6 (cont.): The final uplift interventions that were used in the study. These were updated based on the pilot study as well as the guidelines given in the transformative mediation literature, particularly those given by Jorgensen et al. (2001) for new practitioners of transformative mediation. Based on the pilot study, short verbal utterances were inserted before certain nonverbal interventions to attract participant attention. Also, verbal interventions were introduced earlier to help to frame nonverbal interventions.

Final Uplift Intervention Descriptions and Rationale	
Robotic Interventions	Rationale for Intervention
$\beta_{uplift,4}$ {<Nonverbal cues> [Nonverbal Cue Order] [N1] Wait until there is a break in the discussion. [N2] Display Sad/Angry Face <Verbal cues> [Verbal Cue Order] [V1] "I've sensed some uncertainty/frustration in the discussion. Is that correct?"}	This is a simplification of the verbal interventions used during the pilot study. The robot makes explicit that there is negative emotion in the interaction. It does not assign the negative emotion to either party, so either party is able to acknowledge it. The robot frames its intervention as a question so that the parties can correct it/express how they truly feel. The robot uses a nonverbal display with the verbal display to give context to the nonverbal interventions in the future.
$\beta_{uplift,5}$ {<Nonverbal cues> [Nonverbal Cue Order] [N1] Display Sad/Angry Face [N2] Embody Shameful/Frustrated Bodily Pose [N3] Display Statically For 10 Seconds.}	This nonverbal intervention amplifies the negative emotion in the discussion to show understanding and to help the dyad members acknowledge and deal with that negative emotion directly which is empowering.
$\beta_{uplift,6}$ {<Verbal cues> [Verbal Cue Order] [V1] "Ahem" <Nonverbal cues> [Nonverbal Cue Order] [N1] Display Sad/Angry Face [N2] Embody Shameful/Frustrated Bodily Pose [N3] Display Statically For 10 Seconds.}	This is the same as the previous intervention, but it uses the verbal utterance to draw attention to itself. The dyad can choose how to address that the emotions may or may not be present during the discussion.

Table 6.7: The final encourage empathy interventions that were used in the study. These were updated based on the pilot as well as the guidelines given in the transformative mediation literature, particularly those given by Jorgensen et al. (2001) for new practitioners of transformative mediation. Based on the pilot study, short verbal utterances were inserted before certain nonverbal interventions to attract participant attention. Also, verbal interventions were introduced earlier to help to frame nonverbal interventions.

Final Encourage Empathy Interventions and Rationale	
Robotic Interventions	Rationale for Intervention
$\beta_{empathy,1}$ {<Nonverbal cues> [Nonverbal Cue Order] [N1] Attend to the speaking party. [N2] Turn toward the listening party and back to the speaker every seven seconds. [N3] Attend to the other party when speaker changes. [N4] Turn toward the listening party and back to the speaker every seven seconds.}	The robot treats each individual the same. Turning between the parties frames the interaction as a discussion between the two and acts as a reminder to the speaker to work with the person who is listening.
$\beta_{empathy,2}$ {<Verbal cues> [Verbal Cue Order] [V1] "Ahem" <Nonverbal cues> [Nonverbal Cue Order] [N1] Attend to the speaking party. [N2] Turn toward the listening party and back to the speaker every seven seconds. [N3] Attend to the other party when speaker changes. [N4] Turn toward the listening party and back to the speaker every seven seconds.}	There is a verbal "ahem" added for the robot to draw attention to itself, so the parties are alerted to the attentive and now responsive agent.
$\beta_{empathy,3}$ {<Nonverbal cues> [Nonverbal Cue Order] [N1] Wait until there is a break in the discussion. [N2] Smile <Verbal cues> [Verbal Cue Order] [V1] "I'm enjoying the chat you two are having"}	This is an explicit framing of the discussion as a conversation between the two people. It also acts to empower the dyad.

Table 6.7 (cont.): The final encourage empathy interventions that were used in the study. These were updated based on the pilot as well as the guidelines given in the transformative mediation literature, particularly those given by Jorgensen et al. (2001) for new practitioners of transformative mediation. Based on the pilot study, short verbal utterances were inserted before certain nonverbal interventions to attract participant attention. Also, verbal interventions were introduced earlier to help to frame nonverbal interventions.

Final Encourage Empathy Interventions and Rationale	
Robotic Interventions	Rationale for Intervention
$\beta_{empathy,4}$ {<Nonverbal cues> [Nonverbal Cue Order] [N1] Wait until there is a break in the discussion. [N2] Display Sad Face <Verbal cues> [Verbal Cue Order] [V1] "I've sensed some discomfort during the discussion. Is that correct?"}	This is a simplification of the verbal interventions used during the pilot study. The robot makes explicit that there is negative emotion in the interaction. It does not assign the negative emotion to either party, so either party is able to acknowledge it. This provides an opportunity for the alienated individual show recognition or look for the negative emotion in future exchanges with her partner. The robot frames its intervention as a question so that the parties can correct it/express how they truly feel, which can be empowering. The robot uses a nonverbal display with the verbal display to give context to the nonverbal interventions in the future.
$\beta_{empathy,5}$ {<Nonverbal cues> [Nonverbal Cue Order] [N1] Attend to the speaking party. [N2] Turn toward the listening party and back to the speaker every seven seconds. When turning toward the listening party, the robot makes a sad face. [N3] Attend to the other party when speaker changes. [N4] Turn toward the listening party and back to the speaker every seven seconds. When turning toward the listening party, the robot makes a sad face.}	This intervention is combining the first intervention with a nonverbal display to amplify the negative emotion in the discussion. It tries to ensure the dyad members are attentive to one another and attentive to the emotions of one another. The explicit acknowledgement of and dealing with negative emotions can be empowering for both members.

Table 6.7 (cont.): The final encourage empathy interventions that were used in the study. These were updated based on the pilot as well as the guidelines given in the transformative mediation literature, particularly those given by Jorgensen et al. (2001) for new practitioners of transformative mediation. Based on the pilot study, short verbal utterances were inserted before certain nonverbal interventions to attract participant attention. Also, verbal interventions were introduced earlier to help to frame nonverbal interventions.

Encourage Empathy Interventions and Rationale	
Robotic Interventions	Rationale for Intervention
$\beta_{empathy,6}$ {<Verbal cues> [Verbal Cue Order] [V1] "Ahem" <Nonverbal cues> [Nonverbal Cue Order] [N1] Attend to the speaking party. [N2] Turn toward the listening party and back to the speaker every seven seconds. When turning toward the listening party, the robot makes a sad face. [N3] Attend to the other party when speaker changes. [N4] Turn toward the listening party and back to the speaker every seven seconds. When turning toward the listening party, the robot makes a sad face.}	This briefly draws attention to the robot, so the dyad members are aware of and have the opportunity to be responsive to the intervention.

6.3 Participant Overview

There were 24 participants run as part of this study. There were 12 participants who had the higher-power role in the interaction and 12 participants who had the lower-power role in the interaction with an intervening robot present. The results from these participants are compared to the 28 participants from the study presented in chapter five in which the unengaged robot was present. There were 14 lower-power and 14 higher-power participants tested with the unengaged robot present.

The demographic information for all participants is summarized in Tables 6.8 to Table 6.12 and Figures 6.3 to 6.7. There were predominately males in all of the four groups (Table 6.8 and Figure 6.3). The groups were composed entirely of Georgia Tech students, largely undergraduates, with the average age of all groups falling near 21 (Table 6.9 and Figure 6.4). The majority of participants in all groups entered the discussion against the new concealed carry law; that is, they believed that concealed carry should be banned on the Georgia Tech campus (Table 6.10 and Figure 6.5). All of the groups followed a unimodal distribution with respect to how much they engaged with the news (Table 6.11 and Figure 6.6). Participants in all groups kept up with the news to some extent, but few actively engaged others in discussions about current events. Finally, participants on the whole knew very little about artificial intelligence. In all four groups, the vast majority of participants did not have any formal training in artificial intelligence (Tables 6.12 and Figure 6.7).

Table 6.8: All of the groups had majority of male participants.

Breakdown of Participants' Genders for Each Study Group			
Condition	Male	Female	Nonbinary
Unengaged Robot, Lower-Power Participant	11	3	0
Unengaged Robot, Higher-Power Participant	10	4	0
Intervening Robot, Lower-Power Participant	6	5	1
Intervening Robot, Higher-Power Participant	7	5	0

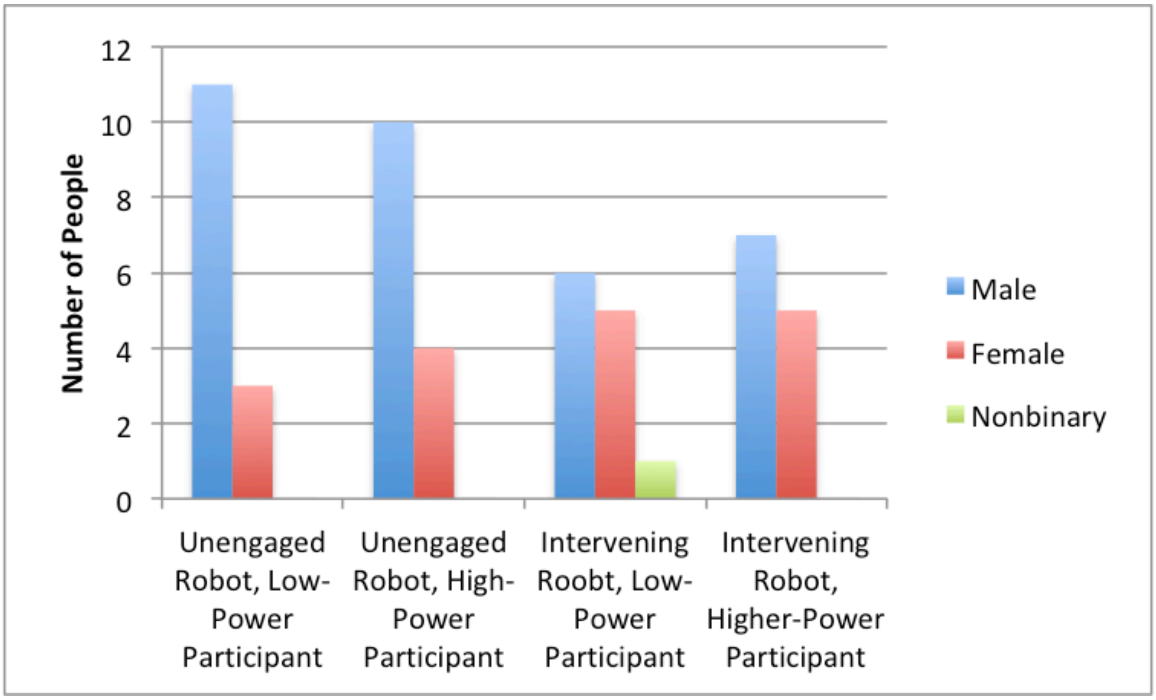


Figure 6.3: There were a total of 34 males included as a part of the study, 17 females, and one nonbinary participant. Each group was composed of mostly male participants.

Table 6.9: The age range for the study was 18-32. The mean age and standard deviation is shown for each group.

Summary of Participants' Ages for Each Study Group		
Condition	Mean Age	Standard Deviation
Unengaged Robot, Lower-Power Participant (n = 14)	21.57	4.309
Unengaged Robot, Higher-Power Participant (n = 14)	20.93	2.556
Intervening Robot, Lower-Power Participant (n = 12)	21.50	3.477
Intervening Robot, Higher-Power Participant (n = 12)	20.58	2.843

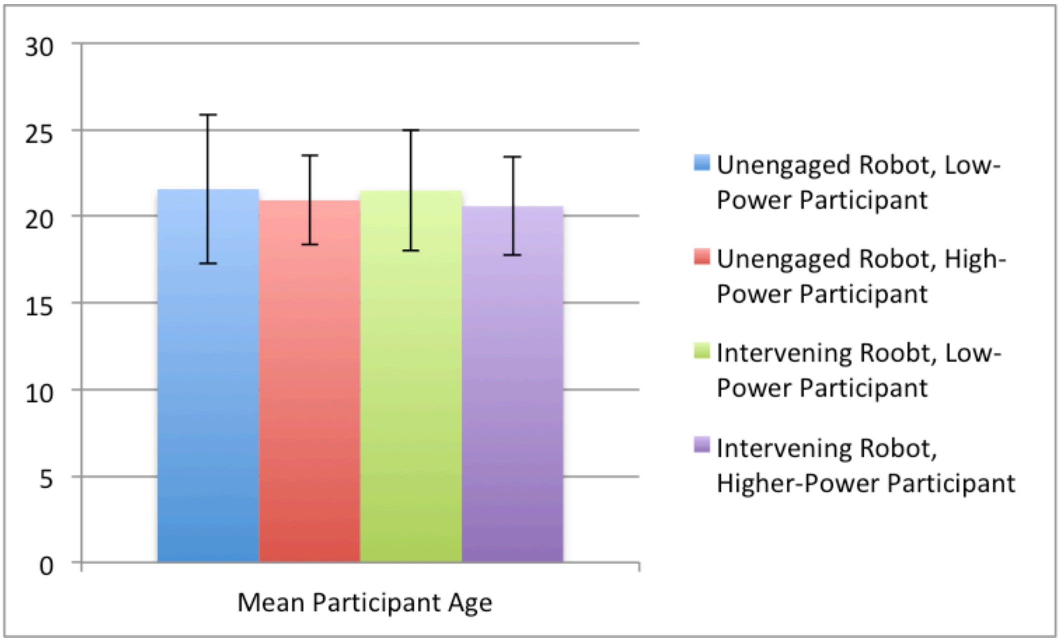


Figure 6.4: The participants were all Georgia Tech students who had a strong opinion on concealed carry. The majority of students were undergraduates. The average age for all groups fell around 21. There was not a significant difference between groups in age.

Table 6.10: The majority of participants in all groups felt less safe with the new concealed carry law. They felt concealed carry on the campus should be banned.

Breakdown of Participants' Opinions on the Discussion Topic for Each Group		
Condition	Against Conceal Carry	For Conceal Carry
Unengaged Robot, Lower-Power Participant	12	2
Unengaged Robot, Higher-Power Participant	12	2
Intervening Robot, Lower-Power Participant	9	3
Intervening Robot, Higher-Power Participant	7	5

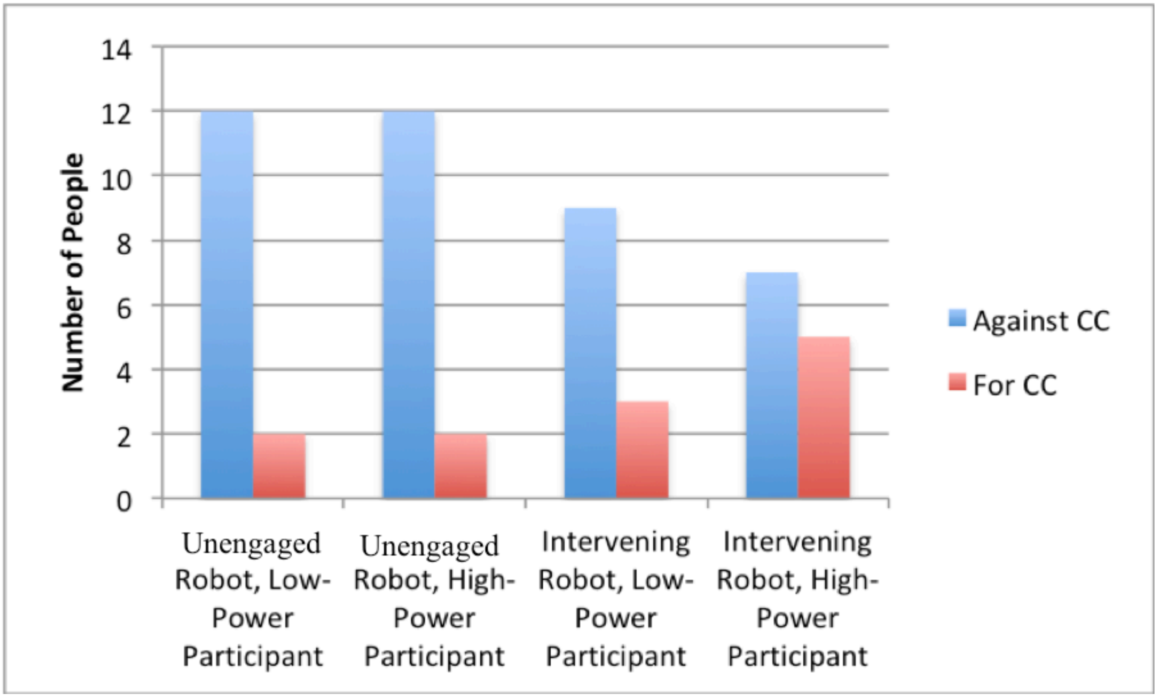


Figure 6.5: Participants had to indicate whether they felt less safe with concealed carry on Georgia Tech or as safe/safer with concealed carry being allowed. The majority of participants in all groups felt less safe and wanted the new law repealed.

Table 6.11: In all four conditions, most participants engaged with news sources, but they did not actively discuss current events with others on a regular basis.

Breakdown of Participants' Media Engagement				
Condition	I don't keep a close eye on current events.	I watch or read the news once in a while.	I watch or read from a variety of news sources regularly.	I watch or read from a variety of news sources regularly and actively engage others in discussion about various issues.
Unengaged Robot, Lower-Power Participant	1	6	5	2
Unengaged Robot, Higher-Power Participant	1	5	6	2
Intervening Robot, Lower-Power Participant	1	6	3	2
Intervening Robot, Higher-Power Participant	1	5	5	1

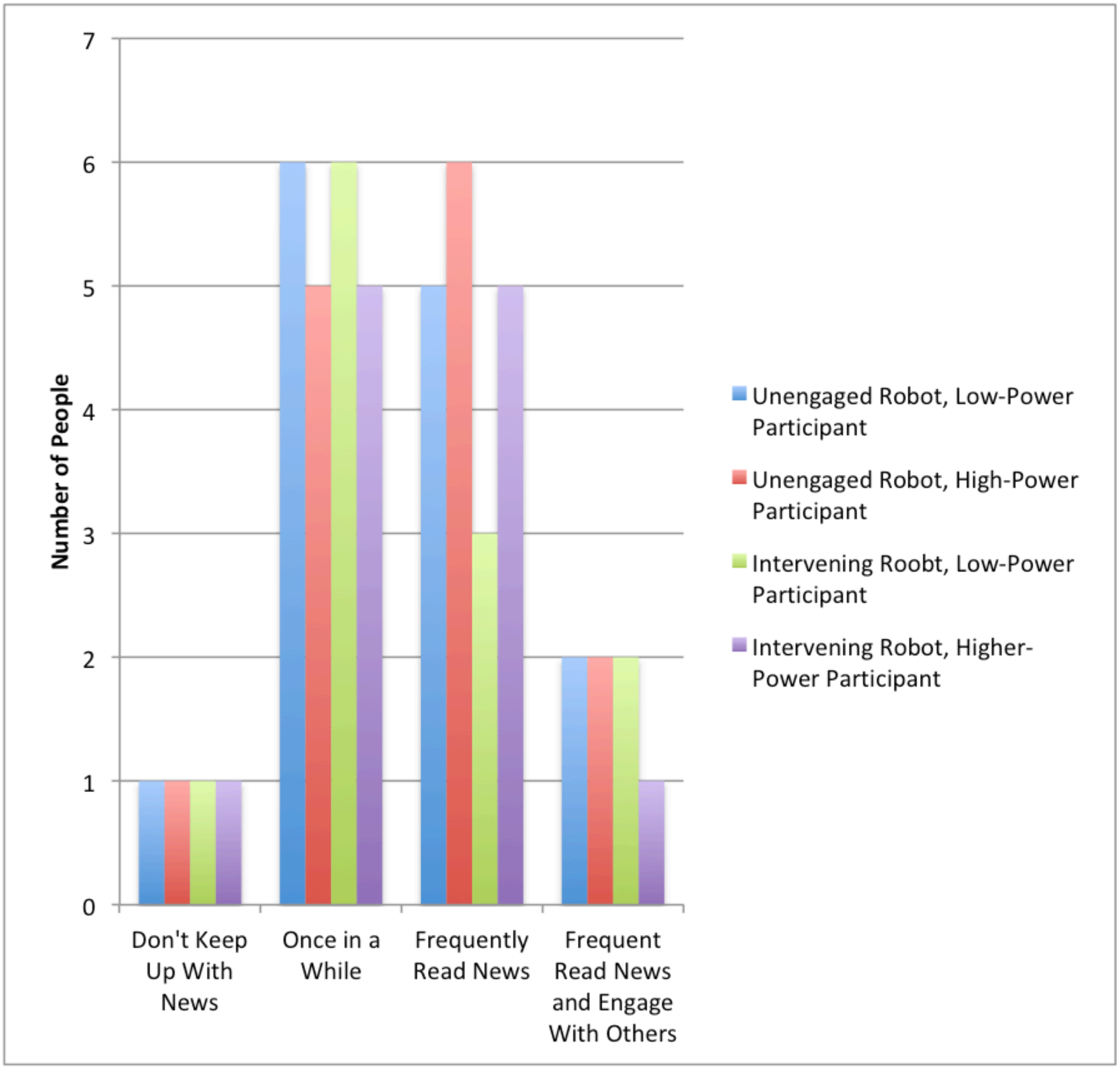


Figure 6.6: All groups followed a unimodal distribution with respect to the participants' engagement with news and current events. The participants in all groups, on average, stayed up-to-date on current events, but they were not regularly engaging others about these events and having discussions.

Table 6.12: The majority of participants in all of the conditions had no formal artificial intelligence education or training. This table gives the number of participants in each condition for each level of artificial intelligence expertise.

Breakdown of Participants' Backgrounds in AI.				
Condition	Exposed to AI in Pop Culture	Had Done Casual Independent Study of AI	Had Taken University Level Classes in AI	AI Researcher
Unengaged Robot, Lower-Power Participant	8	6	0	0
Unengaged Robot, Higher-Power Participant	9	3	2	0
Intervening Robot, Lower-Power Participant	4	7	1	0
Intervening Robot, Higher-Power Participant	7	4	1	0

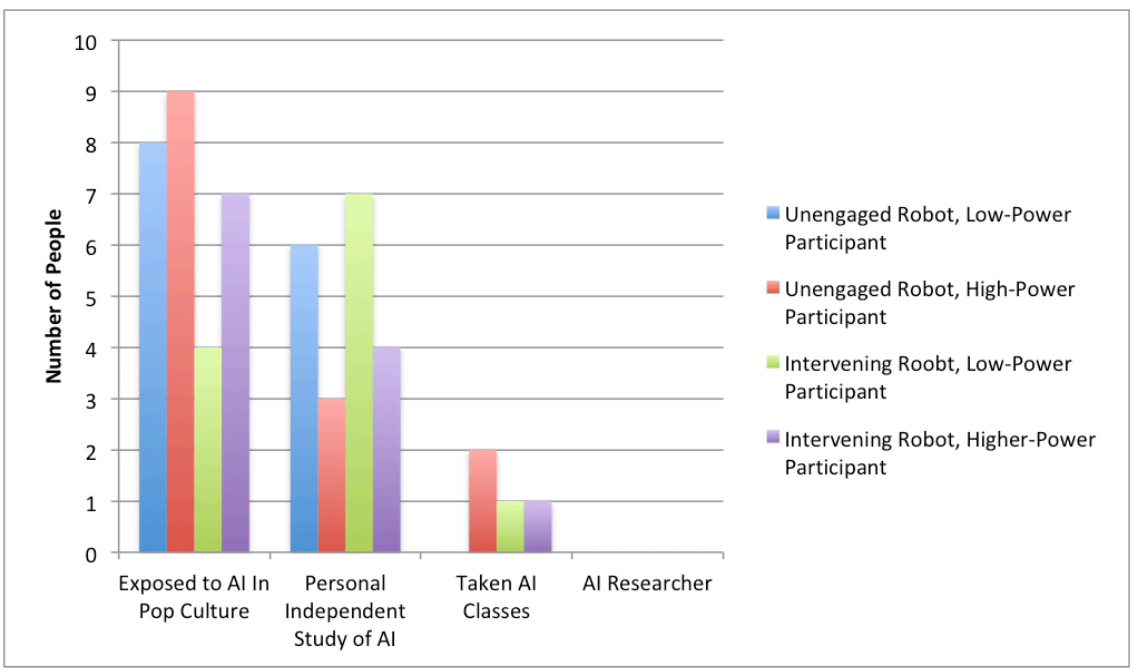


Figure 6.7: The majority of participants in all conditions had only been exposed to artificially intelligent agents in pop culture or through casual independent study.

In addition to completing a demographics survey at the outset of the interaction, the participants completed measures relating to their trait dominance, their opinions on human-human conflict, and their willingness to let a robot support human-human interactions. The results of these measures are summarized in Tables 6.13-6.16 and Figures 6.8-6.11. These measures were gathered before the discussion.

The participants did not have strong opinions with respect to their tendency to lead discussions or go along with what other people wanted to do. The averages for each group were 4 (neutral) or 5 (slightly agree) on these respective questions. They indicated that they were generally comfortable expressing themselves during discussions on contentious topics. The averages for each group were 5 (slightly agree) and 6 (agree) on a scale that ranged from 1 (strongly disagree) to 7 (strongly agree). There were not significant differences ($p > .05$) between groups (see Table 6.13 and Figure 6.9).

The participants were unsure that people in positions of power or those with strong opinions on certain on topics are biased when interacting with people with whom they disagree (Table 6.12 and Figure 6.10). The groups did agree that sometimes two people who disagree need a third-party to help them reach a resolution (Table 6.13 and Figure 6.11).

Subjects were unsure about how the presence of a robot would make them feel compared with that of a third-party person. They also were unsure if a robot ever could capture how they were feeling. Participants, however, were open to a robot trying to support human-human interactions. The average for all the groups was 3 (slightly disagree) when asked if they would feel judged by a robot reflecting their feelings or if they felt a robot could not do anything to help people relate (Table 6.14 and Figure 6.12).

Table 6.13: The participants completed Likert-style measures ranging from 1 (strongly disagree) to 7 (strongly agree) related to their trait dominance. In this case, 4 indicated neutral/unsure. These measures were gathered before the experiment's discussion.

Summary of Participants' Trait Dominance		
Condition	Mean	Standard Deviation
	Responses to: I like to lead when having a group discussion.	
Unengaged Robot, Low Power Participant (n = 14)	4.57	1.4
Unengaged Robot, High Power Participant (n = 14)	4.5	1.16
Intervening Robot, Low Power Participant (n = 12)	4.33	1.78
Intervening Robot, High Power Participant (n = 12)	5.41	1.08
	Responses to: I go along with other people's ideas when deciding what to do.	
Unengaged Robot, Low Power Participant (n = 14)	3.92	1.14
Unengaged Robot, High Power Participant (n = 14)	4.14	1.41
Intervening Robot, Low Power Participant (n = 12)	3.58	1.24
Intervening Robot, High Power Participant (n = 12)	4.167	1.4
	Responses to: I am comfortable expressing my opinions on a contentious topic.	
Unengaged Robot, Low Power Participant (n = 14)	5.643	0.63
Unengaged Robot, High Power Participant (n = 14)	5.286	1.49
Intervening Robot, Low Power Participant (n =12)	5.583	1.44
Intervening Robot, High Power Participant (n =12)	5.083	0.9
	Responses to: I am usually confident in how I express my ideas and myself.	
Unengaged Robot, Low Power Participant (n = 14)	5.92	0.92
Unengaged Robot, High Power Participant (n = 14)	5.36	1.08
Intervening Robot, Low Power Participant (n = 12)	5.33	1.37
Intervening Robot, High Power Participant (n =12)	5.58	1.0

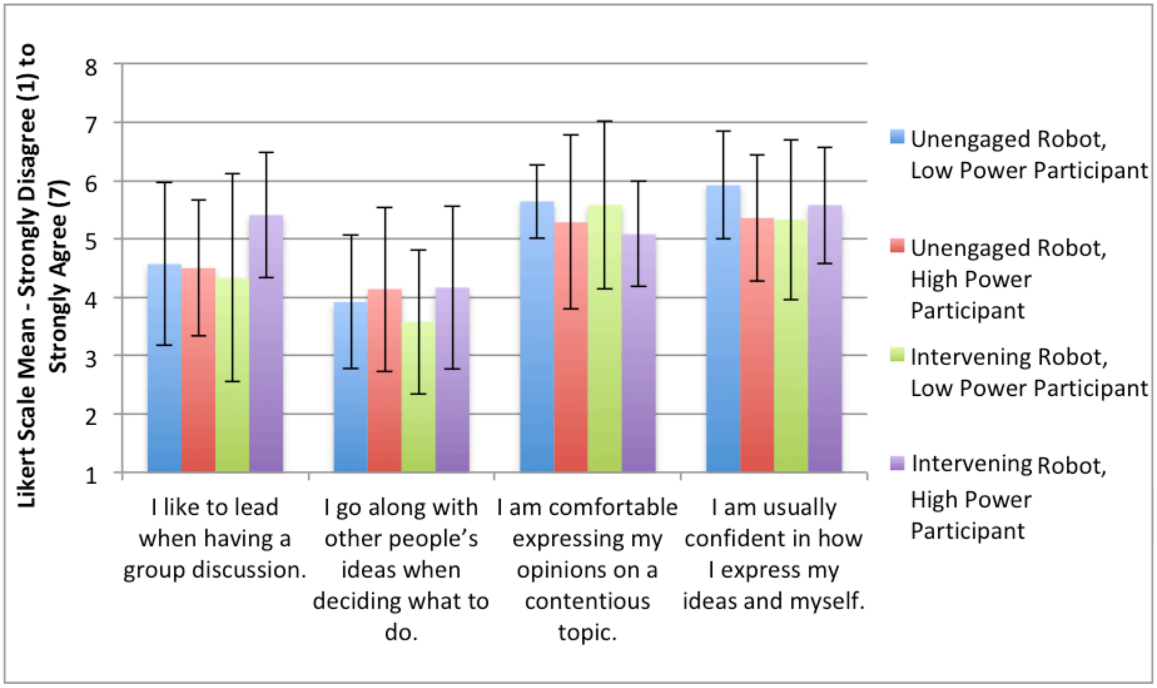


Figure 6.8: Subjects did not have strong opinions with respect to their desire to lead discussions or their tendency to go along with others ideas when making decisions. They did indicate some confidence (5 = slightly agree) in how they expressed themselves and their willingness to engage on contentious topics. These measures were gathered before the experiment's discussion on concealed carry.

Table 6.14: The participants completed Likert-style measures where they gave their opinions on human-human conflict in hierarchical relationships. On these measures, 1 signified strong disagreement, 4 signified being neutral and 7 indicated strong agreement. These measures were gathered before the experiment's discussion on concealed carry.

Participants' Opinions on Hierarchical Human-Human Conflict		
Condition	Mean	Standard Deviation
	Responses to: When I disagree with a superior, I often feel obligated to put their needs and wants above my own.	
Unengaged Robot, Low Power Participant (n = 14)	4.0	1.24
Unengaged Robot, High Power Participant (n = 14)	4.64	1.28
Intervening Robot, Low Power Participant (n = 12)	4.417	1.78
Intervening Robot, Higher Power Participant (n = 12)	3.5	1.45

Table 6.14 (cont.): The participants completed Likert-style measures where they gave their opinions on human-human conflict. On these measures, 1 signified strong disagreement, 4 signified neutrality and 7 indicated strong agreement. These measures were gathered before the experiment's discussion on concealed carry.

Participants' Opinions on Hierarchical Human-Human Conflict		
Condition	Mean	Standard Deviation
	Responses to: When I openly disagree with someone who has more control than I do in a situation, I often feel like my ideas aren't seriously considered.	
Unengaged Robot, Low Power Participant (n = 14)	4.429	1.34
Unengaged Robot, High Power Participant (n = 14)	4.857	1.17
Intervening Robot, Low Power Participant (n = 12)	4.667	1.61
Intervening Robot, Higher Power Participant (n = 12)	4.75	1.36
	Responses to: When I have strong feelings about a topic or decision, it can be difficult to listen to and appreciate alternative viewpoints.	
Unengaged Robot, Low Power Participant (n = 14)	4.071	1.54
Unengaged Robot, High Power Participant (n = 14)	3.357	1.5
Intervening Robot, Low Power Participant (n = 12)	3.917	1.83
Intervening Robot, High Power Participant (n = 12)	3.667	2.15

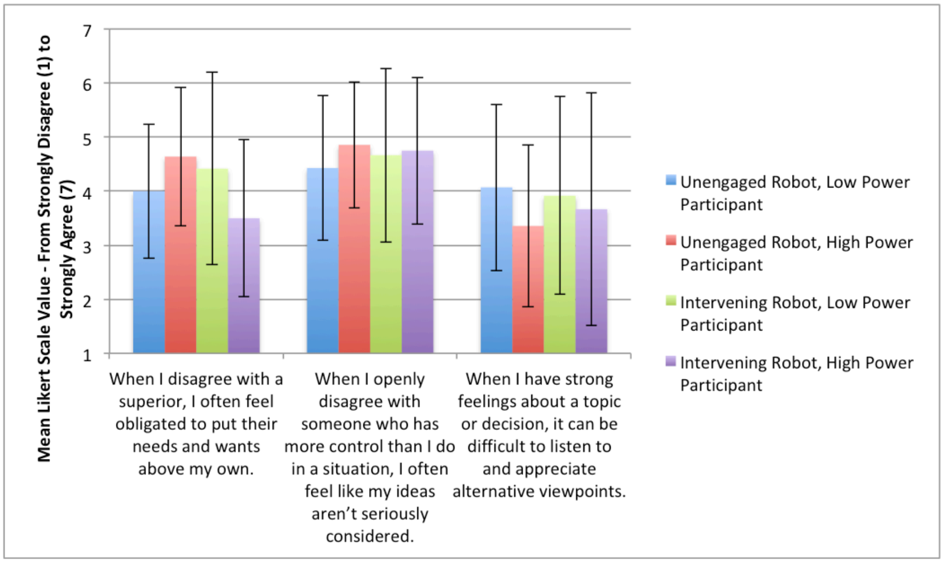


Figure 6.9: Participants in all groups were unsure (4) if the needs of higher-power individuals are favored over those of the lower-power individuals. They were also unsure about their ability to recognize viewpoints other than their own. These measures were gathered before the experiment's discussion on concealed carry.

Table 6.15: All of the groups agreed (on a Likert-style item) that sometimes two people who are involved in a disagreement (conflict) need a neutral third party to help them reach a resolution (5 = slightly agree, 6 = agree). This measure were gathered before the experiment's discussion on concealed carry.

Participants' Opinions on Third-Party Mediation		
Condition	Mean	Standard Deviation
	Response to: Sometimes two people who disagree need a neutral third-party to help them listen to one another and come to mutually satisfactory resolution.	
Unengaged Robot, Low Power Participant (n = 14)	5.857	0.95
Unengaged Robot, High Power Participant (n = 14)	5.42	1.09
Intervening Robot, Low Power Participant (n = 12)	5.67	1.37
Intervening Robot, High Power Participant (n = 12)	6.5	0.52

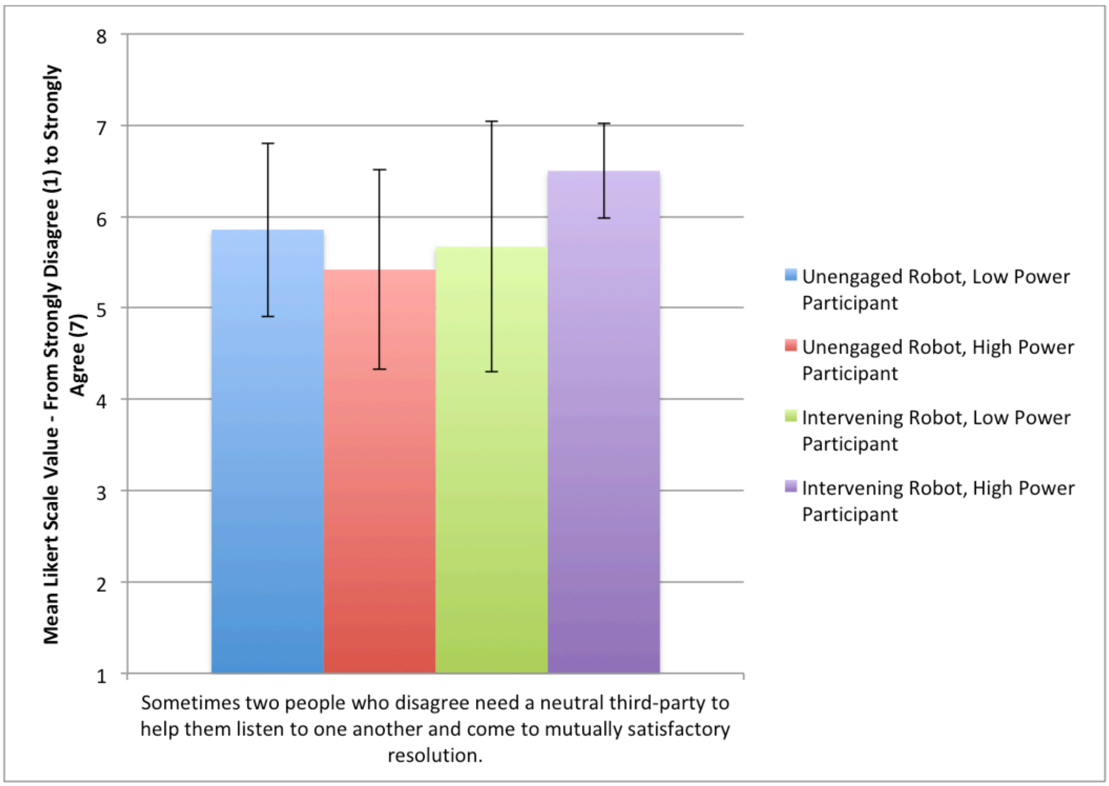


Figure 6.10: The groups agreed that sometimes a neutral third-party is necessary to help two people resolve a conflict. The group with the participant in a higher-power role and an intervening robot present was significantly higher than the other groups ($p < .05$). This measure was gathered before the experiment's discussion on concealed carry.

Table 6.16: The participants completed Likert-style measures ranging from 1 (strongly disagree) to 7 (strongly agree) about a robot's potential to aid human-human interactions. These measures were gathered before the experiment's discussion.

Condition	Mean	Standard Deviation
	Responses to: If I were having a dispute with someone, a robot's presence would make me more uncomfortable than a third person.	
Unengaged Robot, Low Power Participant (n = 14)	3.43	1.5
Unengaged Robot, High Power Participant (n = 14)	3.92	1.21
Intervening Robot, Low Power Participant (n =12)	2.667	1.07
Intervening Robot, High Power Participant (n = 12)	3.75	1.66
	Responses to: A robot could never accurately understand and reflect how I'm feeling or how I'm behaving.	
Unengaged Robot, Low Power Participant (n = 14)	4.5	1.09
Unengaged Robot, High Power Participant (n = 14)	3.7	1.59
Intervening Robot, Low Power Participant (n = 12)	3.33	1.56
Intervening Robot, High Power Participant (n = 12)	3.08	1.44
	Responses to: If a robot were to try to reflect my feelings, I would feel like I was being judged.	
Unengaged Robot, Low Power Participant (n = 14)	2.857	0.86
Unengaged Robot, High Power Participant (n = 14)	2.769	1.09
Intervening Robot, Low Power Participant (n = 12)	2.2917	1.08
Intervening Robot, High Power Participant (n = 12)	3.25	1.36
	Responses to: A robot couldn't do anything that would help me open up to or better identify with another person.	
Unengaged Robot, Low Power Participant (n = 14)	2.857	0.86
Unengaged Robot, High Power Participant (n = 14)	2.769	1.09
Intervening Robot, Low Power Participant (n = 12)	2.917	1.08
Intervening Robot, High Power Participant (n = 12)	3.25	1.36

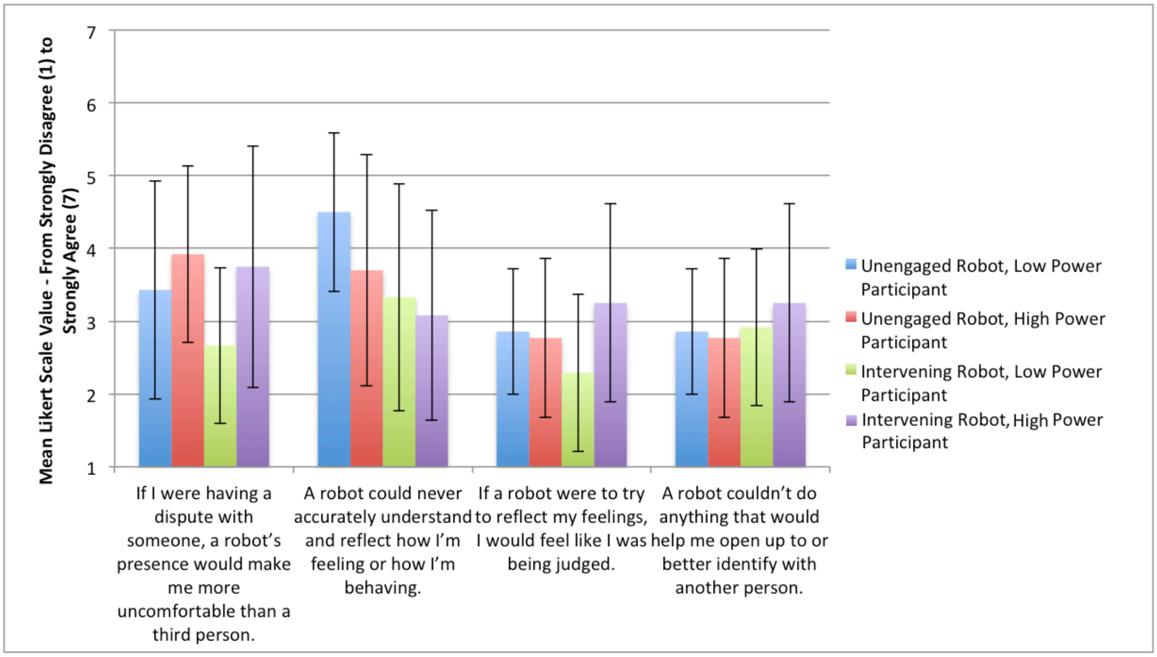


Figure 6.11: The participants were unsure (4) about a robot's presence making them more uncomfortable than a third person. They also were unsure about a robot being able to reflect how they were feeling. Participants, however, were open to a robot trying to reflect their feelings (3 = they slightly disagreed that they would feel judged by the robot reflecting their feelings), and they felt that a robot might be able to do something to help human-human communication (3 = they disagreed that a robot could not do anything to help two people relate). The participants with the intervening robot may have been more willing to believe that the robot. They disagreed more strongly ($p < .05$) with the statement that robot could never accurately understand and reflect their feeling than participants in the unengaged conditions. These were gathered before the experiment's discussion on concealed carry.

6.4 Study Results

The measures used for this study were identical to those used for the study discussed in chapter five and were summarized in full in Table 5.11. To compare the four groups on each of the measures introduced in the previous chapter, we used a two-way ANOVA. There are two factors dividing the groups: power (the participant had a higher- or lower-status) and the behavior of the robot (the robot was unengaged or intervening). This is the standard way to evaluate groups that are divided by more than one factor.

On certain measures, the normality assumption of this test did not hold. There were groups in which the outcome measure (dependent variable) was decidedly not normal. Participants responses sometimes had very little variance or were bimodally distributed at the extremes of the scales. Normality was assessed using the standard Shapiro and Wilk's (1965) test. Previous related research has used nonparametric methods when the normality assumption does not hold (e.g. Jung, 2016). The nonparametric tests for multifactor designs, however, are not standardized (Feys, 2016) and multifactor designs were not used in the comparable papers (e.g. Hoffman et al., 2015; Jung et al., 2015; Jung, 2016).

Therefore, when there was not a significant difference between the groups with higher- and lower-power participant groups on a measure in chapter five's study, the higher- and lower-power participant groups were combined, and an intervening robot group was compared to unengaged robot group on that measure. This was only done for outcome measures where the power-level factor did not have an influence. In the case there was a difference in chapter five's study on a certain measure, multiple Mann-Whitney U Tests (Mann & Whitney, 1947) were used with correction on that measure (Miller, 1966).

To check the internal consistency of scales containing multiple items measuring a single construct, we used Cronbach's α (Cronbach, 1951). If the value fell below 0.7, we tested each item independently.

The first set of outcome measures relate to participants' perceptions of the robot. Participants gave the intervening robot more attention but did not find it disruptive to the conversation. An important finding if a robot is going to support

human-human conversation. Participants did not view the intervening robot as an ally. The robot's interventions were designed to make it clear it was a neutral agent.

Participants completed Likert-style measures ranging from 1 (strongly disagree) to 5 (strongly agree) on which they indicated how much they agreed or disagreed with a certain item. Though neither group agreed that the robot was disruptive, participants with the intervening robot found the robot to be significantly more disruptive than the unengaged robot (Mann-Whitney U Test, unengaged median = 2.0, intervening median = 2.5, $U = 225.5$, $p = .035$). Participants with an unengaged robot present disagreed (2 = disagree) that the robot was disruptive; the participants with the intervening robot were more neutral (3 = neutral) about the statement the robot was disruptive (see Table 6.17 and Figure 6.12).

Objective results lend support to the participants' self-report results. There was not a significant difference ($p > .05$) between the intervening and unengaged robot groups with respect to number of glances participants took at the robot. There was, however, a significant difference (Mann-Whitney U Test, unengaged median = intervening median = 0.0, $U = 210$, $p = .0004$) between groups on the number of times the robot was mentioned by participants. Participants explicitly mentioned the intervening robot more than the unengaged robot (see Tables 6.18 and 6.19 and Figure 6.13). These mentions often took place during or immediately following interventions.

The subjects with the intervening robot had stronger opinions about the robot's ability to understand how they felt during the discussion than those in the unengaged robot condition. In the unengaged robot condition, twenty-three of twenty-seven participants were unsure (3 = unsure) of whether or not the robot understood how they felt. In the intervening condition, there was a wider distribution of responses (see Table 6.20 and Figure 6.14). When comparing the magnitude of the difference between the neutral response and a participant's responses, there was a significant difference (Mann-Whitney U Test, unengaged median = 0.0, intervening median = 1.0, U = 186, p = .002) between the unengaged robot group and the intervening robot group (see Table 6.21). It will be important to understand with future research, including longitudinal studies, how these differing opinions influence the use of similar technologies.

Finally, subjects in both groups were largely neutral (3 = neutral) when asked about whether or not the robot was on their side during the discussion (see Table 6.22 and Figure 6.15). The intervening robot group, however, disagreed more strongly with the notion that the robot was on their side. There was a significant difference between groups (unengaged median = intervening median = 3.0, U = 231, p = .02) on this measure (see Table 6.23). The interventions used by the robot treat both the participant and the confederate the same and make it clear that the robot is not on either individual's side.

Table 6.17: Participants completed a Likert-style item as to whether or not the robot was disruptive. They disagreed the unengaged robot was disruptive (2 = disagree). The participants with the intervening robot present were more neutral (3 = neutral). This measure was gathered after the study's discussion.

Participants' Impressions of Whether or Not Robot was Distracting			
Condition	Response to Prompt: The robot was disruptive.		
	Total Number of Participants	Median	Standard Deviation
Intervening Robot	24	2.5	1.167
Unengaged Robot	28	2	1.054

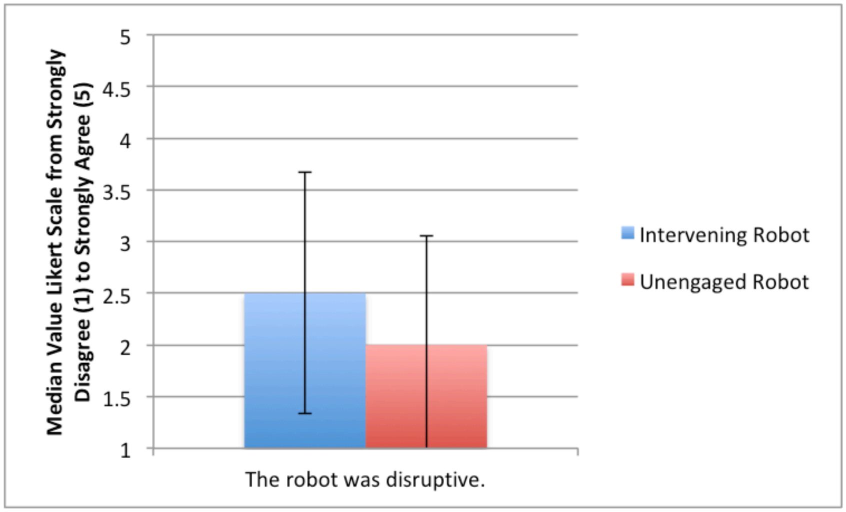


Figure 6.12: Participants answered a Likert-style item about the robot being disruptive. The intervening robot was rated as significantly more disruptive ($p < .05$) than the unengaged robot. This measure was gathered after the study's discussion.

Table 6.18: The unengaged robot was not mentioned explicitly by any of the participants. There were 9 participants who referenced the intervening robot.

Count of Participants Who Mentioned and Did Not Mention Robot		
Condition	Did Not Mention The Robot	Mentioned The Robot
Intervening Robot	15	9
Unengaged Robot	28	0

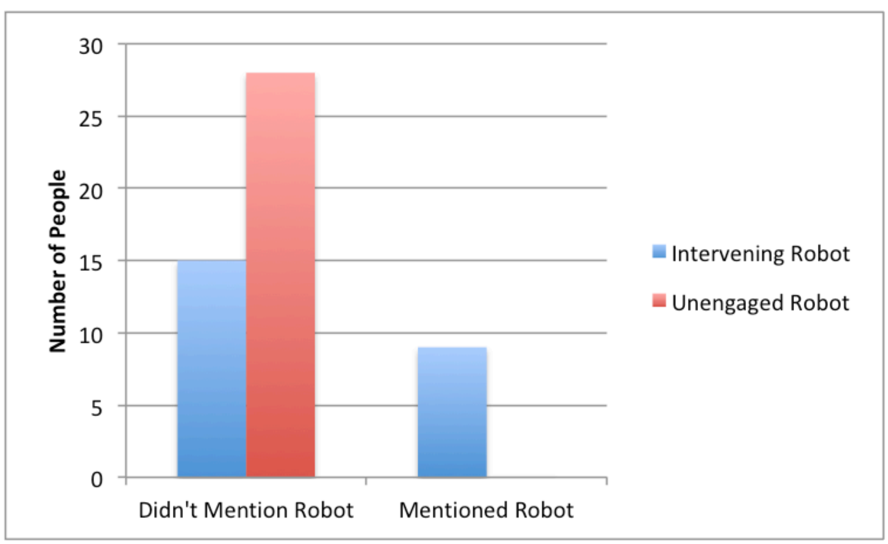


Figure 6.13: None of the participants mentioned the unengaged robot. There were some participants who mentioned the intervening robot. These mentions came during or immediately following interventions.

Table 6.19: This is the average and median number of time participants explicitly mentioned the robot during the discussion. There were more explicit mentions of the robot in the intervening robot condition when compared with the unengaged condition. The interventions drew attention to the robot. The intervening and unengaged groups were significantly different with respect to making explicit mention of the robot ($p < .05$).

Summary of the Number of Times the Robot was Mentioned			
Condition	Average	Median	Standard Deviation
Intervening Robot (n = 24)	.5	0	.78
Unengaged Robot (n = 28)	0	0	0

Table 6.20: There is a clear difference in the distribution in the responses received when participants were asked if the robot understood how they felt during the discussion. The intervening robot had a wider distribution of responses than the unengaged robot. This measure was collected after the study's discussion

Breakdown of Participants' Impressions of the Robot's Situational Understanding					
Condition	Responses to the Prompt: The robot clearly didn't understand how I felt.				
	Number of People Who Responded that they Strongly Disagree	Number of People Who Responded that they Disagree	Number of People Responded that they Were Neutral	Number of People Who Responded that they Agree	Number of People Who Responded that they Strongly Agree
Intervening Robot	2	7	10	5	0
Unengaged Robot	0	2	23	1	1

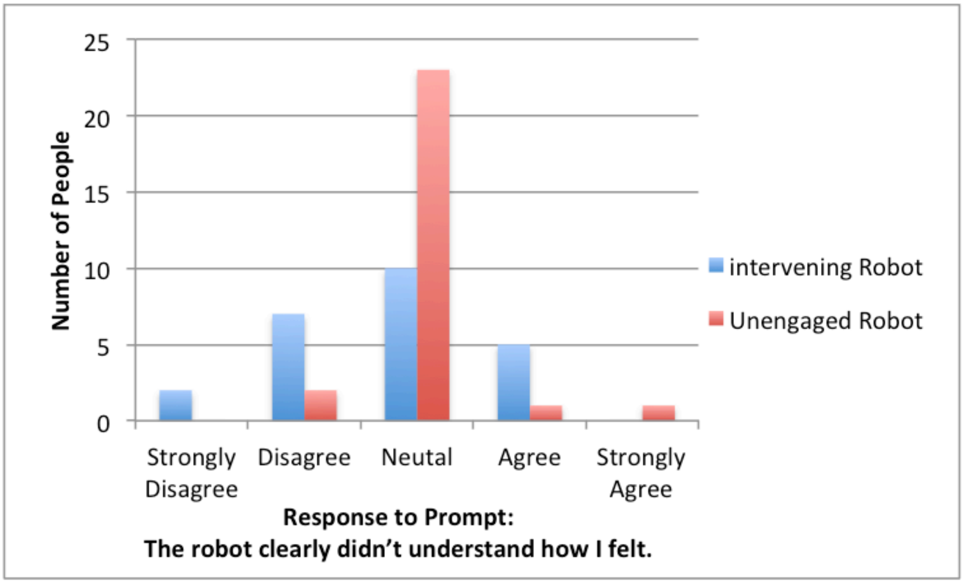


Figure 6.14: Only four out of twenty-seven participants in the unengaged robot condition agreed or disagreed that the robot understood how they felt during the discussion. The rest were neutral/unsure. The intervening robot condition had a much wider distribution. This measure was collected after the study's discussion.

Table 6.21: The absolute difference between each participant's response and the neutral response was taken for the item: "The robot clearly didn't understand how I felt". There was a significant difference between participants with the intervening robot compared to the unengaged robot ($p < .05$). The participants with the intervening robot were less neutral in their impressions regarding the robots understanding of how they felt. This measure was gathered after the study's discussion.

Average Participants' Views of Robot's Situational Understanding			
Condition	The magnitude of the difference between an individual's response and the neutral response to the prompt: The robot clearly didn't understand how I felt.		
	Number of People	Median	Standard Deviation
Intervening Robot	24	1.0	.637
Unengaged Robot	27	0.0	.483

Table 6.22: Again there was a clear difference in the distribution of responses with respect to the item: "The robot was on my side". Participants with the unengaged robot were almost exclusively neutral to this statement (25 of 27 responded neutrally). The intervening robot's participants varied in their responses (with far more people disagreeing 9 compared to 1). This measure was gathered after the study's discussion.

Breakdown of Participants' Beliefs that Robot was on Their Side					
Condition	Responses to the Prompt: The robot was on my side.				
	Number of People Who Responded that they Strongly Disagree	Number of People Who Responded that they Disagree	Number of People Responded that they Were Neutral	Number of People Who Responded that they Agree	Number of People Who Responded that they Strongly Agree
Intervening Robot	4	5	13	2	0
Unengaged Robot	0	1	25	1	0

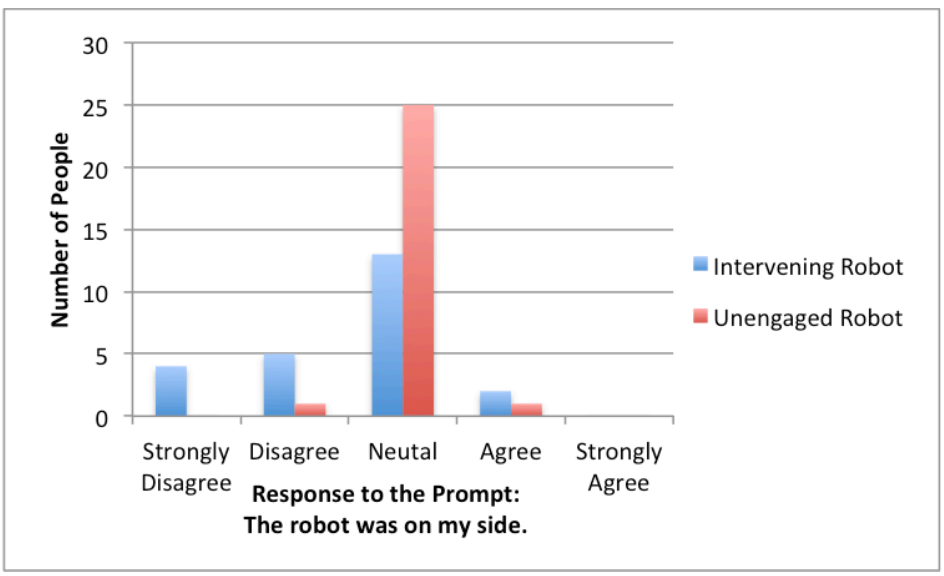


Figure 6.15: Twenty-five of twenty-seven participants in the unengaged condition indicated that they were neutral toward the statement that the robot was on their side. The participants with the intervening robot were more divided. The majority (thirteen) indicated being neutral; however, nine indicated disagreement and a total of eleven were not neutral. This measure was gathered after the study's discussion.

Table 6.23: The individuals with the intervening robot disagreed (2 = disagree) with the statement, "The robot was on my side", significantly more ($p < .05$) than participants in the unengaged robot condition (3 = neutral). The interventions made it clear that the robot was not an ally to the participant. It treated both participant and confederate the same. This measure was gathered after the study's discussion.

Participants' Views About the Robot Being on Their Side				
Condition	Response to Prompt: The robot was on my side.			
	Total Number of Participants	Mean	Median	Standard Deviation
Intervening Robot	24	2.542	3.0	.884
Unengaged Robot	27	3.0	3.0	.277

The section immediately below highlights how the robotic interventions were not able to change the amount of weakness and alienation displayed by an average participant. This is a negative result. The cycle of weakness and alienation that defines destructive conflict was present as much with the interventions as without the interventions. The discussions with interventions had as much negative emotion as those without. This does not mean, however, the agent had no value. The important differences it did make are described subsequently.

Neither the power level of the participant nor the robotic interventions influenced the conflict dynamics. Participants in the unengaged and intervening robot groups rated the conflict dynamics in their discussions with the experiment confederate similarly. There were no significant differences ($p > .05$) between groups on Jehn's (1995; 2001) emotional- and task-conflict scales. Emotional conflict was rated as low in all of the groups with the averages falling between 1 and 2 (on a scale of 5). Task-based conflict was rated to be more moderate in both groups with average between 3 and 4 (on a scale of 5). There were also no significant differences ($p > .05$) between the groups on the empowerment and recognition items (averages between 4 and 5 on all items). In both groups, the individuals reported feeling empowered and reported that they gave and received recognition from the experiment confederate.

The experimenter and a video coder rated segments of video for each experiment participant just as they did for the study described in chapter five. The two indicated whether the subject showed weakness or empowerment in each segment as well as whether the subject gave or refused recognition in each speaking segment. The rating guidelines used by the video raters appear in Appendix A. The agreement between raters was $\kappa = .512$ on the weakness ratings. The agreement between raters was $\kappa = .643$ on recognition ratings. This is moderate agreement (Simand & Wright, 2005) and is comparable to the agreement by coders in previous studies (e.g. Jung, 2016).

The proportion of video segments showing weakness and alienation in each of the four groups was compared using a two-way ANOVA. There was not a significant difference between each of the four groups ($p > .05$) and no interaction effects were found between the groups (see Table 6.24 and Figure 6.16). In all four groups, the video raters identified a moderate amount of weakness and alienation. Neither the interventions of the robot nor the power of the participant altered the conflict dynamics. In the groups, between three and five out of every ten rated segmented showed weakness or alienation.

Table 6.24: The average participant in both conditions showed moderate weakness. In all groups the mean proportion of video segments showing weakness was above .3. **This means over three out of every ten video segments showed some weakness for the average participant.** Similarly, the mean proportion of video segments showing a lack of recognition was above .4. **For the average participant, more than four out of every ten rated video segments showed a lack of recognition.**

Proportions of Video Segments Showing Weakness and Alienation			
Condition	Number of People	Mean	Standard Deviation
	Proportion of Video Segments that showed Weakness		
Unengaged Robot Low Power Participant	14	.402	.249
Unengaged Robot High Power Participant	14	.492	.255
Intervening Robot Low Power Participant	12	.425	.174
Intervening Robot High Power Participant	12	.313	.272
	Proportion of Statements that showed a Lack of Recognition		
Unengaged Robot Low Power Participant	14	.496	.237
Unengaged Robot High Power Participant	14	.515	.259
Intervening Robot Low Power Participant	12	.516	.246
Intervening Robot High Power Participant	12	.421	.359

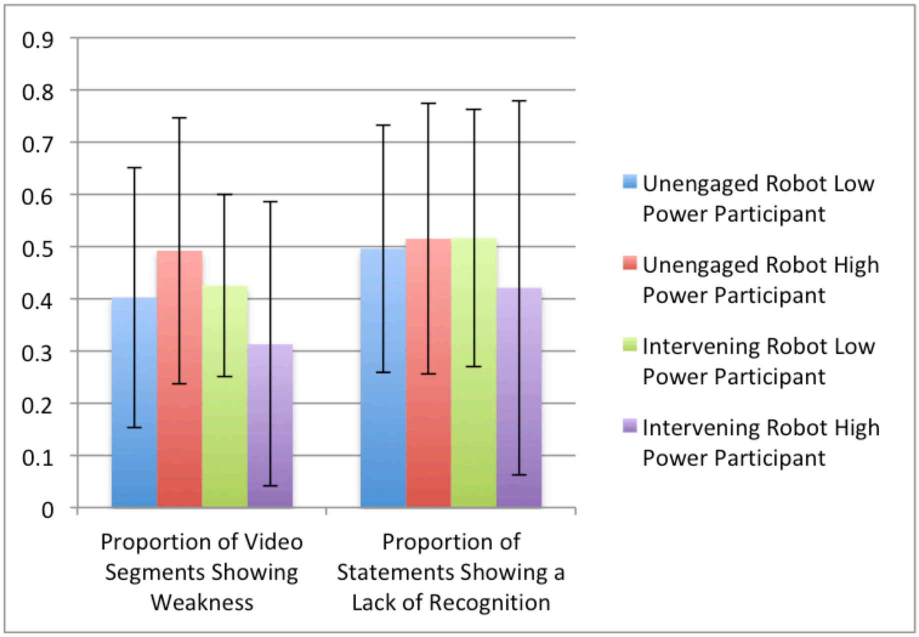


Figure 6.16: There was no significant differences (p > .05) between the groups with respect to the proportion of video segments showing weakness nor with respect to the proportion of video segments showing alienation. **This is an indication that neither the power difference nor the robotic interventions changed the amount of weakness and alienation seen during the discussions.**

Though the conflict dynamics did not seem to be different between groups, there were changes in the participants' communication patterns with the confederate. The way in which people communicate can make them more likely to move out of the destructive cycle of weakness and alienation. Specifically, speaking openly about emotion and speaking clearly about one's thoughts can provide more opportunity for reconnection (Bush & Folger, 2010a; Retzinger, 1991). The intervening robot seemed to encourage participants to communicate in this way more often than the unengaged robot.

There were significantly more explicit mentions of emotion in the intervening robot group than in the unengaged robot group (Mann-Whitney U Test, unengaged median = intervening median = 0.0, U = 196, p = .0002). Participants with the unengaged robot did not explicitly mention their emotional state during discussions with the confederate. There were participants in the intervening robot condition who mentioned that they were unsure of themselves, a little bit uncomfortable, and/or struggling to accurately articulate themselves (see Table 6.25 and Figure 6.17).

Anytime a person is experiencing or showing weakness, there is an opportunity for that individual to move or shift to a stronger position (e.g. Bush & Folger, 2010a). When someone refuses to recognize the viewpoints of the person with whom she is interacting, she can have a recognition shift by subsequently understanding and/or considering the positions of the other. When participants in the study showed weakness during a video segment, there was an opportunity for that participant to have an empowerment shift in the next video segment. When they refused to recognize the confederate's views in a video segment, they had the opportunity for a recognition shift in the next video segment.

The proportion of successful recognition shifts to potential recognition shifts was not significantly different ($p > .05$) between the intervening robot group and the unengaged robot group. The proportion of successful shifts of empowerment to the potential shifts of empowerment was significantly different (Mann-Whitney U Test, unengaged median = .279, intervening median = .5, U = 134, p = .047). The participants with the intervening robot showed moves of empowerment a higher proportion of the time than participants in the unengaged robot condition (see Table 6.26 and Figure 6.18).

Table 6.25: The experimenter coded for explicit mentions of emotion by study participants. Participants in the unengaged robot condition did not mention how they were feeling in the moment (any discomfort or uncertainty). Participants in the intervening condition did mention their feelings in the moment. **The intervening robot made participants more open regarding their current emotional state, which is critical to a successful conflict process.**

Measure	Intervening Robot	Unengaged Robot
Mean Number of Times Feelings Mentioned	.542	0
Median Times Feelings Mentioned	0	0
Standard Deviation	0.72	0
Total Number of Participants	24	28
Number of Participants Who Did Not Mention Feelings	14	28
Number of People Who Did Mention Their Feelings	10	0

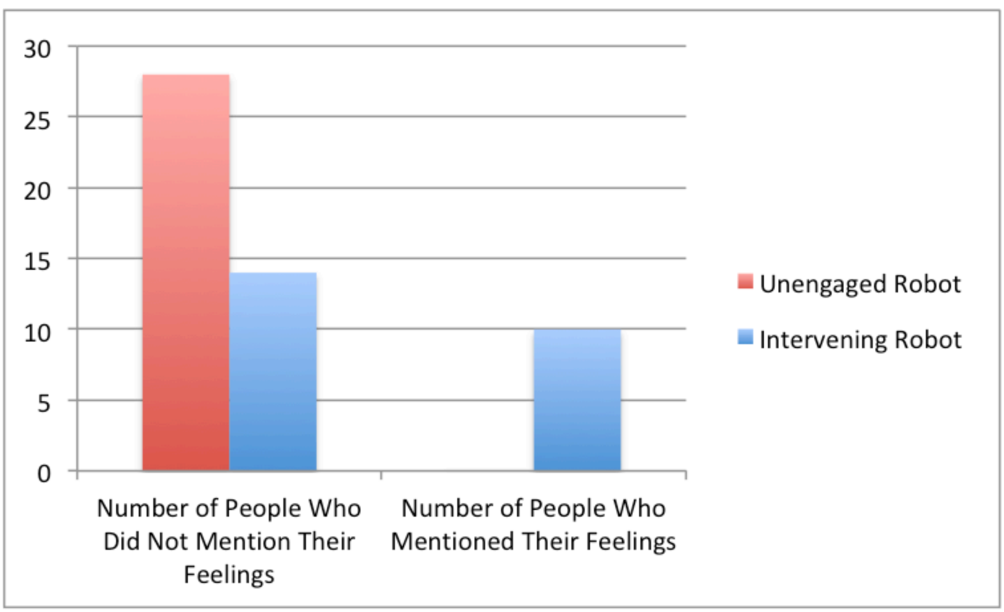


Figure 6.17: Ten out of twenty-four participants explicitly mentioned negative feelings during the discussion (uncertainty, discomfort, etc.). None of the participants in the unengaged robot group explicitly mentioned negative feelings.

Table 6.26: A move of empowerment is a when a participant demonstrated a lack of empowerment in a coded video segment before demonstrating emerging empowerment in a subsequent video segment. Therefore, every video segment where a lack of empowerment occurs is an opportunity for a move of empowerment (in the following video segment). The proportion of successful moves of empowerment for the participants in the unengaged and intervening robot groups are summarized. The proportion of successful recognition shifts are also given. These are calculated in an identical way to the moves of empowerment. A move of recognition is when genuine recognition follows a refusal of recognition. **In the intervening robot condition, participants more often made moves of empowerment following demonstrations of weakness. This is critical for moving out of a destructive conflict cycle to a constructive conflict cycle.**

Proportion of Video Segments Showing Moves of Empowerment and Recognition				
Measure	Participant Group	Number of Participants	Median	Standard Deviation
Proportion of Moves of Empowerment to Opportunities for Empowerment	Intervening Robot	19	.5	.377
	Unengaged Robot	22	.279	.375
Proportion of Moves of Recognition to Opportunities for Recognition	Intervening Robot	19	.333	.396
	Unengaged Robot	25	.333	.414

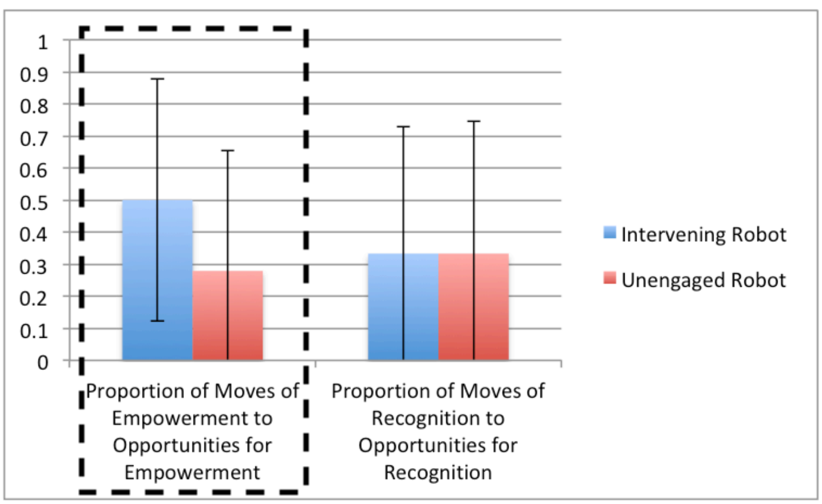


Figure 6.18: In the intervening robot condition, there was a significantly higher ($p < .05$) proportion of moves of empowerment than in the unengaged robot condition. That means when an individual demonstrated weakness in one video segment, they had **a higher likelihood of demonstrating empowerment in the next video segment in the intervening condition compared with the unengaged condition.**

The participants' perceptions of the experiment's confederate depended on their power level as well as the behavior of the robot. A two-way ANOVA was used to examine the interaction between the robot's behavior and the power-level of the participant on how engaged and warm the confederate was perceived to be (see Table 6.27 and Figure 6.19). There was no interaction effects ($F(1, 51) = 3.45$, $p = .069$); however, the test showed that lower-power participants viewed the confederate as significantly warmer and more engaged than higher-power participants ($F(1, 51) = 8.9$, $p = .005$). There was no main effect for the behavior of the robot ($p > .05$).

Two Mann-Whitney U tests were run to examine the influence of the participant's power level and the robot's behavior on the perceived dominance of the experiment's confederate (see Table 6.28 and Figure 6.20). The participant's power level influenced the perceived dominance of the confederate; higher-power participants disagreed significantly more with the statement that the confederate "dominated the conversation" (Mann-Whitney U Test, higher-power median = 2.0, lower-power median = 3.5, $U = 203.5$, $p = .011 < .025$).

A Mann-Whitney test was run to compare the differences between the group of participants with the intervening robot and the unengaged robot present with respect to the perceived inclusivity of the confederate (see Table 6.29 and Figure 6.21). Participants with the unengaged robot present found the confederate to be more inclusive (promote similarity and have conversations with depth) than participants with intervening robot present (Mann-Whiney U Test, unengaged robot median = 4.577, intervening robot median = 4.077, $U = 481.0$, $p = .008$).

Table 6.27: Participants in the higher-power position viewed the experiment confederate as less warm and engaged than participants in the lower-power role. This agrees with what was found in chapter five.

Condition	**Global Intimacy Measure – Involvement/Affection**		
	Number of People	Mean	Standard Deviation
Unengaged Robot Low Power Participant	14	5.621	0.638
Unengaged Robot High Power Participant	14	4.813	0.379
Intervening Robot Low Power Participant	12	5.192	0.56
Intervening Robot High Power Participant	12	5.026	0.655

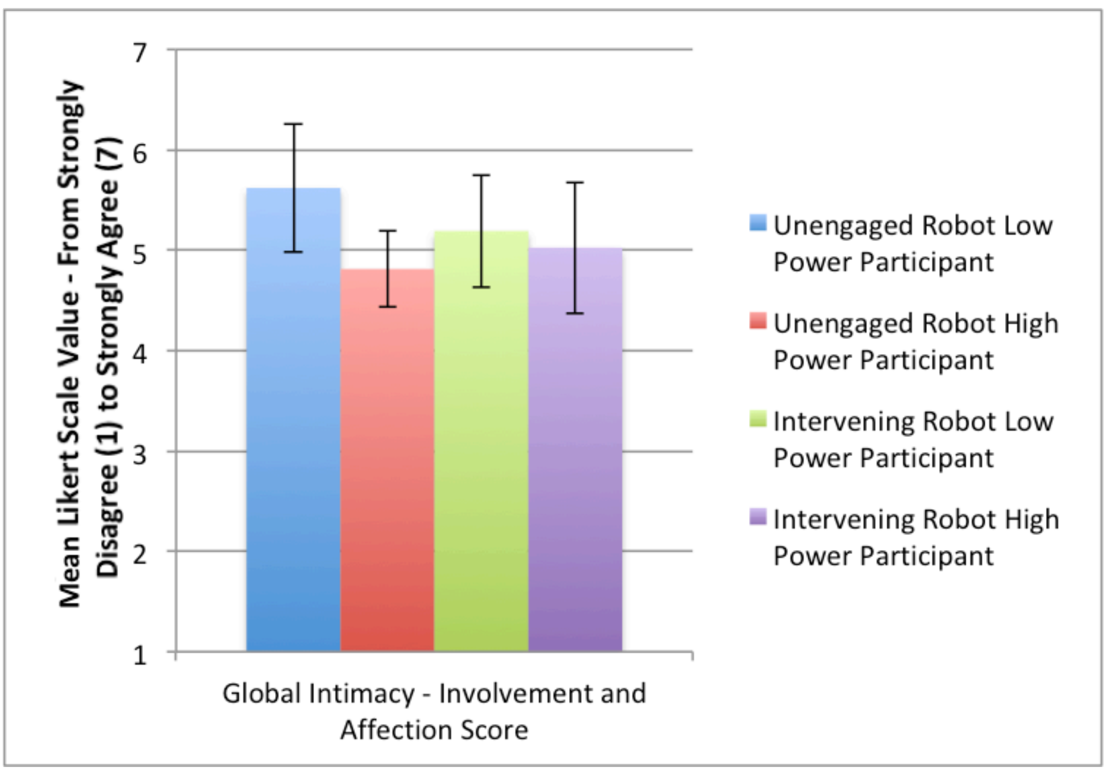


Figure 6.19: Participants in the higher-power conditions saw the experiment confederate as significantly colder and detached ($p < .05$). There were no differences with respect to the actions of the robot ($p > .05$). There were also no interaction effects ($p > .05$).

Table 6.28: The higher-power participants disagreed significantly more ($p < .025$) with the statement that their partner (the experiment confederate) had dominated the conversation. This agrees with what was found in the experiment presented in chapter five.

Summary of Likert-style Measure Related to Partner Dominance			
Condition	Number of Participants	Median Value	Standard Deviation
	Responses to: Your partner dominated the conversation.		
Lower Power Participants	26	3.5	1.16
Higher Power Participants	26	2.0	1.0

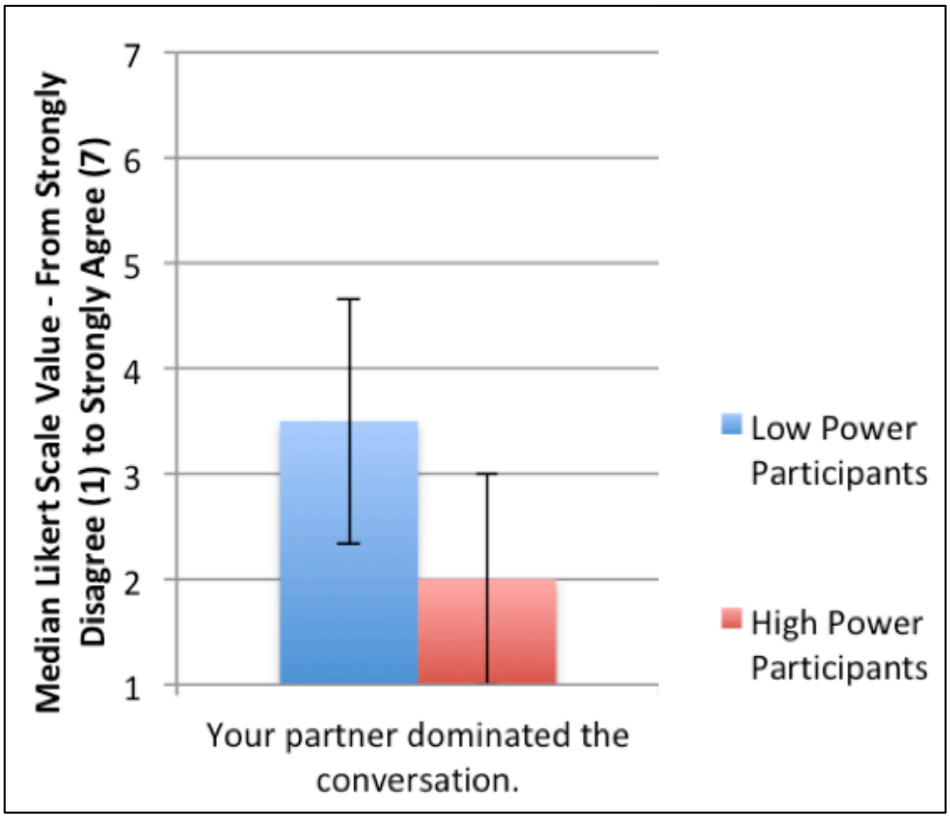


Figure 6.20: The lower-power participants viewed the experiment confederate as more dominate during the discussion. This supports that the experiment manipulation to create a power-difference in the relationship was successful.

Table 6.29: **Participants with the intervening robot present viewed the experiment confederate as significantly less inclusive** ($p < .05$) during the discussion than participants with the unengaged robot. The experiment confederate made the participants feel as though they had less in common with one and less as though they were having conversations of depth. **The intervening robot highlighted the confederate's perpetual disagreement.**

Summary of Likert-style Measure Related to Partner Inclusiveness			
Condition	Number of Participants	Median Value	Standard Deviation
Unengaged Robot	28	4.577	0.756
Intervening Robot	24	4.077	0.710

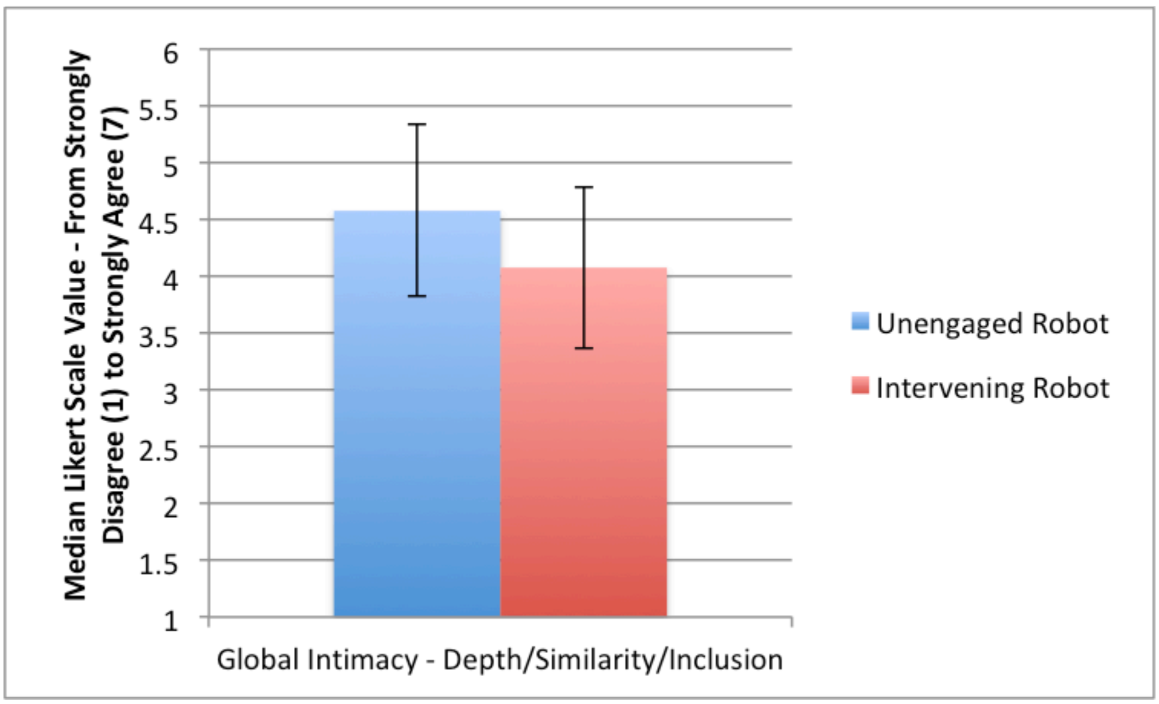


Figure 6.21: The intervening robot caused the participants to perceive a greater separation between the themselves and the study's confederate. **The robot highlighted the confederate's continued rebuke of every participant idea by the confederate.**

Finally, the experimenter and video coder from above rated video segments for a specific states of weakness (negativity) and a specific state of alienation (insensitivity). The negativity state (as described in section 4.1) is a state of weakness where a lower-power participant pushes away or withdraws from the higher-power confederate. Insensitivity (as described in section 4.1) is a state of alienation in which the higher-power participant is not responsive to or mocking of the lower-power confederate. Each video segment was rated as active or inactive as described in Appendix B. The agreement between the raters on negativity was $\kappa = .556$. On insensitivity, the rate of agreement was $\kappa = .597$. This is moderate agreement (Simand & Wright, 2005) and is comparable to the agreement by coders in previous studies (e.g. Jung, 2016).

The autonomous system contained a while loop that checked whether or not the negativity or insensitivity was active or inactive during trials for the entire discussion (see section 4.2). This was so that the robot was able to intervene when the algorithm believed there was a problematic state active in the relationship. If the algorithm labelled the state as active at any time during a particular video segment, that segment was labelled as active. If the algorithm did not label the video segment as active, then the video segment was considered to be labelled inactive by the autonomous system. The ground truth that this label is compared against are the labels given by the video coders. The video segments on which the video coders agreed were the only segments compared. The results for the autonomous system are contained in Table 6.30.

As described in chapter five, there are certain individuals who use these cues to convey weakness and alienation. **There are times in which these cues will be completely absent and weakness and alienation are present**. Further, these cues can mean very different things in the larger context (Bush, 2010; Retzinger, 1991). **These cues only provide moderate positive predictive values when present.** These cues are a slice of the interaction. **A more complete context can improve identification of these**

states (Bush, 2010; Noce, 2010). As certain technologies mature, such as natural language processing technologies, a more complete context can be achieved.

Table 6.30: The precision, recall and specificity for the autonomous robotic algorithm during the discussions with the intervening robot present for negativity (a state of weakness) and insensitivity (a state of alienation). **The nonverbal and paralinguistic cues used by in the model's current implementation have moderate positive predictive power.** Additional cues (such as those derived with natural language processing) to provide a more complete context may aid identification accuracy.

Outcome Measures on Test Data for Computational Model Implementation			
State	Precision	Recall	Specificity
Negativity	.625	.61	.545
Insensitivity	.58	.73	.737

6.5 Study Discussion

This study was about determining if an intervening autonomous robot can support positive changes during conflict in a hierarchical relationship when compared to the presence of an unengaged robot. The robot used in this study modeled the relationship using nonverbal and paralinguistic cues that are commonly associated with weakness and alienation in the transformative mediation and destructive conflict literature (e.g. Bush & Folger, 2010a; Moen et al., 2001; Retzinger, 1991). When it determined there was an active state of weakness and/or alienation in the relationship (as described in section 4.2 of the thesis), it used interventions (described in sections 4.3 and 6.2) rooted in the transformative mediation literature to support empowerment and recognition shifts.

Although the robot did not improve the proportion of weakness and alienation showed by participants, it was able to change the participants' communication in meaningful ways that can support the amelioration of conflict in other relationships. With respect to the participants' perceptions of the conflict, they rated the task-based and emotional-conflict level similarly regardless of whether the intervening or unengaged robot were present. The intervening robot did not change their perceptions of empowerment and alienation according to self-report either. Similarly, there was not a significant difference between the proportion of video segments showing weakness or those showing a refusal of recognition as rated by the video coders when the intervening robot was compared to the unengaged robot.

Objectives measures show the participants with the intervening robot did explicitly express emotion more than the participants with the unengaged robot. Openly acknowledging weakness is the critical action to restore healthy relationship dynamics during conflict (Retzinger, 1991). The participants with the intervening robot also had a higher proportion of video segments that showed emerging empowerment following video segments that showed weakness than the participants with the unengaged robot present. When participants make a move of empowerment, they are more likely to begin a cycle of mutual empowerment and recognition (Bush & Folger, 2010a). Empowerment is often a prerequisite for recognition (Bush, 2010). Therefore, if participants are showing empowerment more frequently following weakness, there is more opportunity to begin a positive conflict cycle to move out of the destructive cycle of weakness and alienation.

The ways in which this study helped to answer each of the subsidiary questions from the first chapter are summarized in Table 6.31. The following chapter is a summary of the work presented in this dissertation. It includes the contributions of the dissertation, the answers to the questions asked as part of this dissertation, as well as the future work for the dissertation. I argue that a robotic agent is able to support positive and meaningful changes in a hierarchical dyad’s conflict.

Table 6.31: Summary of each subsidiary question from the first chapter and how this chapter has helped to answer that question.

Subsidiary Question	Insights Provided by the Study
How is an unengaged robot, that has been introduced as something that is meant to support communication within the relationship, perceived by the dyad members and how do interventions change this perception?	• The relationship-focused interventions made the robot more distracting. o Participants indicated the robot was more distracting according to self-report and explicitly mentioned the robot more. • Participants were more opinionated about the robot's ability to understand their emotions. • The intervening robot was not viewed as not being on the participants side during the discussion.
How can a social robotic agent represent a dyad's relationship state, problematic or otherwise, and decide when to intervene in the relationship?	• The algorithm that identifies the problematic relationship states has moderate positive predictive power. • The behavioral cues used by this model will appear in moments that do not possess relationship strain and be absent in moments in which there is relationship strain (Retzinger, 1991). • As technology allows for the identification of additional cues (such as those derived from natural language), improvements can be made to how the agent decides to intervene (Folger, 2010).

Table 6.31 (cont.): Summary of each subsidiary question from the first chapter and how this chapter has helped to answer that question.

Subsidiary Question	Insights Provided by the Study
What channels of communication should the agent use when supporting the dyad, how overt should these communications be in order to avoid upsetting the relationship, and how should the agent choose between interventions?	• The unengaged robot largely fell into the background in the study presented in chapter five. • Nonverbal interventions were not noted by participants in the pilot study (a small verbal utterance preceded certain nonverbal interventions) • The interventions caused a significant increase in the proportion of video segments showing empowerment following a video segment that showed weakness compared to the robot's unengaged behavior. • Participants rated the intervening robot as more distracting and mentioned it more, but they did not agree that the robot was distracting, nor did they look at it more than the unengaged robot.
What issues does the power dynamic in the relationship present for the robotic agent when trying to provide a conflict process that is viewed as fair and equitable by both participants?	• Higher-power participants viewed the confederate as colder and less dominant than lower-power participants. • Interventions were designed to treat the dyad members the same and be neutral (to make clear to both dyad members that the robot was not on either person's side).

6.6 Study Limitations and Shortcomings

The study presented here has several limitations. This section considers each of these shortcomings in turn. The claims and contributions of the study are considered in light of these limitations and opportunities for future research to overcome these limitations is discussed.

Experimenter Bias

Perhaps the greatest limitation of this study is the fact that an experimenter was involved in the generation of the video coding and objective measures. Appendix E is dedicated to addressing experimenter bias in this study. It reviews the measures that were generated as a part of this study, how they were generated, how the experimenter attempted to minimize the bias introduced into these measures, and the evidence that exists for the reliability of these measures. Because experimenter bias exists for these study measures, claims that rely on these data alone are suggestive. Future research should seek to have several professional relationship-focused mediators who are blind to experiment hypotheses code the gathered data.

Single Robot Morphology

This study only included Robokind's R25 platform. This platform provided an array of verbal and nonverbal channels with which to implement interventions rooted in the transformative mediation literature. Previous to this study there had not been any research implementing relationship-focused interventions on any artificial platform and claims are made with the understanding that an embodied agent was used that had the capabilities to intervene verbally and nonverbally (with bodily and facial gestures). Previous research has shown how morphology and embodiment has an influence on various perceptions of the robot (e.g. persuasiveness, social presence, etc.) (Lee et al. 2006; Li, 2015). Future studies can incorporate additional agents with different embodiments and morphologies to see how employing the principles of transformative mediation with different agents is similar and different to this humanoid agent.

Limited Predictive Power of Relationship Computer Model

As described at length in sections 5.8 and 6.4, the percepts used by the autonomous robotic agent to identify problematic relationship states were limited in their ability to identify this relationship. The computational model showed only moderate positive predictive power. Additional cues, especially verbal cues identified through natural language processing, will need to be incorporated for agents to have comparable perceptive capabilities to human mediators. Future studies can incorporate these verbal cues as natural language processing technologies mature. Future studies with larger datasets may also try to use unsupervised approaches to try to learn additional nonverbal and paralinguistic features that can be employed by computational system. The limitations of the autonomous system's ability to identify problematic relationship states were acknowledged in this experiment and the interventions that were used were rooted in guidelines for newer mediators who sometimes also have trouble identifying the most appropriate moments to intervene (Jorgensen et al., 2001).

Confederate-Participant Relationship

The confederate-participant relationship in this study was a relationship between strangers that would only meet in this single instance. When it comes to supporting relationships between people, the greatest value is seen in supporting relationships that will continue. The way in which participants responded to the study confederate and to the robot were likely different than the ways in which participants would respond if it was someone that they knew they were going to have to interact with again or had interacted with before. Future studies should use a longitudinal approach that inserts agents in real hierarchical dyads to understand if the changes that the robotic interventions made in these relationships will continue over time and whether additional changes occur when participants have more stake in making the relationship work.

Length of Interaction

The limited time in which participants had to interact with the confederate meant that it was difficult to observe substantive and lasting relationship changes. Further, the limited interaction time meant that there was an insufficient time to learn baseline readings for meaningful data like physiological data. Additional time with a robot present in a dyadic relationship would allow for the agent to learn the context of a particular relationship. The agent could learn how the dyad interacts when the relationship is considered to be healthy. Experimenters could also assess if the interventions by the robot help to sustain a healthy relationship over time. This, again, indicates that future studies should take a longitudinal approach in which the agent is embedded into interactions with regularly meeting dyads.

CHAPTER 7

CONCLUSION

At its core, human-human conflict is a state in which two people are disconnected from one another (Bush & Folger, 2010a). This state is painful for both and can create long-term problems if a healthy conflict process is not used to resolve the issues the two are having. Chapters one and two of this thesis describe how unhealthy conflict processes are common in and particularly destructive to hierarchical relationships. They also argue that a robot is uniquely suited to help to support communication in these relationships and to help to ameliorate potentially problematic human-human exchanges. This argument led to our asking the primary research question: *How can social robots be used to support positive change in relationships with power differentials that are experiencing negative conflict?*

To answer this requires answering four different subsidiary questions, which have been addressed throughout the thesis. These subsidiary questions are reexamined in the first subsection below along with the contributions that have been made by the thesis. Subsequently, future research directions and concluding remarks are given.

7.1 Research Questions Revisited and Resulting Contributions

To understand how a robot can have a positive influence in a negative interaction between two people in a hierarchical relationship, one must understand how the robot will be perceived by the dyad. If the robot is overly disruptive or is negatively perceived by either dyad member, regardless of his/her power-level, the dyad may not be responsive to the robot or may actively disuse it. It is also important to know how a robot can identify problematic states in human-human relationships. A robot should intervene to support a relationship if and when that relationship is strained. Finally, it is necessary to answer how the robot should intervene when a relationship is in a negative state. The agent's interventions should be understood by the dyad it is trying to support and interventions should have the desired effect.

Our conclusions for each of the subsidiary questions are summarized below. In addition, the contributions that have been made by the thesis while answering these questions are highlighted at the section's conclusion.

1. *How is an unengaged robot, that has been introduced as something that is meant to support communication within the relationship, perceived by the dyad members and how do interventions change this perception?*

The second chapter (Section 2.3, Table 2.2) of this dissertation reviewed literature related to peoples' perceptions of robotic agents. Researchers have found that the perception of a robot can be strongly influenced by the metaphor a person uses to understand the agent. In the studies covered in chapters five and six, the robot was clearly introduced as a peripheral tool to support communication that would not be actively participating in the discussion. The study in chapter five had an unengaged robot present (the robot moved subtly and in ways not corresponding to the conversation). In this study, the robot faded into the background of the interaction for both the higher- and lower-power participants. It was rated as not distracting on self-report measures, participants explicitly stated on surveys that they forgot the robot was there, and the robot was not mentioned during the discussion (Section 5.6).

The interventions caused the robot to command more attention according to the study presented in chapter six (Table 6.31). According to self-report measures, the intervening robot was significantly more distracting than the unengaged robot, and it was explicitly mentioned more than the unengaged robot (Tables 6.17-18, Figure 6.13-14). The robot, however, was not rated as a "disruption" (according to a Likert-style measure, Table 6.17), and there was not a significant difference in the number of times participants looked at the robot in the intervening and unengaged conditions (Section 6.4). The interventions made it clear the robot was not on the participant's side (Table 6.22-23, Figure 6.16)). As covered in questions 3 and 4 below, interventions were designed to be neutral and tentative.

2. *How can a social robotic agent represent a dyad's relationship state, problematic or otherwise, and decide when to intervene in the relationship?*

Models for human-human communication and human mediation have been researched extensively. The second chapter covers a specific model for destructive human conflict; it is based in the idea that destructive conflict stems from a cycle of mutual weakness (shame) and alienation (hostility) within and between dyad members (Section 2.1). I present a general and novel computational model to identify problematic relationship states in hierarchical relationships that is rooted in this literature (Section 3.1). The states for the model are operationalized using existing work in relationship-focused human-human mediation (Section 4.1).

A complete implementation for the identification of two problematic relationship states is provided (Section 4.2). This implementation uses cues related to shame and anger (e.g. averted gaze, loudness of voice) to identify weakness and alienation in hierarchical human-human relationships. Testing of this implementation, which has moderate positive predictive value, was carried out (Sections 5.8, 6.4), and the thesis affords an analysis of the model's implementation along with guidelines for how future research can improve results (Sections 5.8, 6.4). Additional cues can improve results by providing a more complete situational context as technologies, like natural language processing, mature.

3. *What channels of communication should the agent use when supporting the dyad, how overt should these communications be in order to avoid upsetting the relationship, and how should the agent choose between interventions?*

Transformative mediation, a relationship-focused style of human mediation, was derived from the conflict theory used to help to answer subsidiary question two. It serves as the basis on which this thesis developed relationship-supporting robotic interventions. The background of this mediation technique is covered in chapter two (Section 2.2). A mediator's role in destructive conflict interactions is to help communication between dyad members by supporting and trusting the decision-making of both dyad members. This includes using minimal encouragers to have dyad members speak openly, amplifying dyad members' emotions to allow for mutual recognition, and acting in ways that help the dyad to see the interaction is a discussion between them (not competing monologues). The thesis presents an action-selection mechanism, which chooses a manifestation of a robotic behavior to support relationship communication. These robotic behaviors as well as their manifestations are rooted in the transformative mediation literature (Sections 2.2, 4.3). Complete implementations of relationship-supporting interventions for two robotic behaviors are presented in the thesis's fourth chapter (Section 4.3). The study presented in chapter six (Section 6.4) suggests that these relationship-focused interventions have a positive influence on communication in strained relationships. Our data shows that dyad members explicitly mentioned their feelings more with a robot enacting these interventions when compared to the unengaged robot. They made empowered statements more often following weak statements with the intervening robot when compared to the unengaged robot. Dyad members took note of interventions (Section 6.2); however, participants did not view the robot as a distraction (Section 6.4).

4. *What issues does the power dynamic in the relationship present for the robotic agent when trying to provide a conflict process that is viewed as fair and equitable by both participants?*

Transformative mediation stresses the point of not taking sides during the interaction (Section 2.2). To take sides is to lessen the decision-making power of one of the dyad members, which weakens this dyad member and worsens the destructive conflict cycle. Therefore, the robot's interventions were carefully designed to treat the party's the same (Section 4.3). The study in chapter five showed that a participant's power level in the interaction did not influence how he/she perceived the robot (Section 5.7). Interventions made it clearer that the robot was not on participants' side during the interaction (Section 6.4). The higher- and lower- power dyad members had typical conflict dynamics develop with the robot present (Sections 5.7, 6.4). The power structure created within the dyad at the interaction's outset was maintained. Lower-power participants viewed the confederate as more dominant than higher-power participants after in the interaction.

Contributions

- **A Novel and General Computational Model to Identify Problematic Relationship States in Human-Human Relationships**

  As described in chapters one and two (Section 2.3) of the thesis, robotics researchers have begun to examine how technology can benefit human-human interactions and relationships. The technologies developed up until this point, however, have been responsive to single cues such as shouting (e.g. Hoffman et al., 2015). They have not devised representations of human-human relationships that can help to inform how the relationship can be supported. Depending on single cues to assess interaction health has been shown to be insufficient to understand if the relationship is strained (Bush, 2010; Noce, 2010).

  In the second chapter (Section 2.1), we described how relationship-focused human mediators are able to identify problematic states as third parties during human-human interactions. These mediators identify the alignment of many features within an interaction to identify a certain state of mutual weakness and/or alienation in a relationship between two people that they are then able to respond to. In the third chapter (Section 3.1), we introduced a general computational architecture based on this literature, which allows for a robotic agent to identify states of weakness and alienation.

This model provides the means by which a robotic agent can recognize strain in a human-human relationship by identifying specific states of weakness and alienation so it can meaningfully intervene to support the relationship. This model is general and can be expanded by other researchers to include additional sensors, features, and states if they are exploring other types of relationships or other states of weakness and alienation.

- **Operationalization, Implementation, and Testing of Two Problematic Hierarchical Relationship States on a Robot**

Previous research has not presented algorithms or code to identify states of weakness and alienation in hierarchical human-human relationships. These relationships are some of the most susceptible to have problematic states arise, and problems in these relationships can have long-term consequences for lower-power relationship members (as described in Section 2.1). The computational model introduced in the third chapter of the thesis (Section 3.1) identifies six problematic relationship states in hierarchical relationships. Two of these states are grounded in the literature and fully operationalized with complete algorithms presented in the thesis's fourth chapter (Sections 4.1, Algorithms 4.1-4.4 and Table 4.9). The grounding and implementation of these two states can inform the work of future researchers trying to design functional systems to identify strife in human-human relationships.

Further, the studies in the fifth (Section 5.8) and sixth (Section 6.4) chapters provide an analysis of the model's implementation. This showed the current implementation had moderate precision. The negative relationship states were active approximately half the time it indicated that they were active. This quantifies the value current perception technologies have when it comes to identifying these particular states and reveals shortcomings that future research should address if artificially intelligent systems are to support hierarchical human-human relationships.

- **A Novel and General Action-Selection Mechanism for Relationship-Focused Interventions**

The research into how technology can intervene to support human-human relationships up until this point has presented systems that tightly couple a single behavioral response to a single percept (e.g. Hoffman et al., 2015). As covered in Section 2.2 of the thesis, human mediators have several strategies for relationship-focused interventions, and the ways in which they employ these strategies are very context specific. The action-selection mechanism introduced in Section 3.2 couples robotic behaviors with the problematic relationship states identified in the computational model to allow for the agent to make more targeted relationship interventions (interventions that depend on the relationship's context). The action-selection mechanism selects different manifestations of these behavioral interventions, which allows the agent to have varied responses (responses that try different intervention strategies). This general action-selection mechanism gives a computational framework that can be used by other researchers to design relationship-focused technologies that better mirror human mediators.

- **Implementation and Testing of Robotic Intervention Behaviors Rooted in Human Mediation Theory**

It is not a certainty that robots should support relationships in the same manner as human mediators do. As discussed in the Section 2.3 of the thesis, there are important differences in peoples' perceptions of robots compared to humans, and there are clear differences in the capabilities of the two. Previous research has not connected robotic interventions with the numerous relationship-focused human mediation principles discussed here. This thesis grounded robotic relationship-focused interventions in the principles of a human mediation theory (Section 4.3). The study in chapter six suggests that a robot employing these principles makes a positive difference in a human-human relationship. Specifically, our data suggests that an intervening robot was able to encourage more explicit expressions of feelings and a higher proportion of empowered statements following statements of weakness than a non-intervening robot. Participants with the intervening robot recognized the confederate's lack of responsiveness to their feelings more so than those with the unengaged robot (according to self-report measures) (Section 6.4). Future research into relationship-supportive technologies can build on work done by human experts. This thesis has provided evidence that suggests at least certain principles put forth by human mediation experts can be applied to implementations on a robotic platform with success.

7.2 Future Directions

The experimental results gathered in the thesis offer three clear directions through which the work can be extended. Our data suggests that an intervening robotic agent is able to increase the explicit expression of feelings and the proportion of empowered statements a participant makes when compared to an unengaged robotic agent. The next step could be to compare the intervening robotic agent to a human intervening in similar ways or artificial agents of other embodiments intervening in similar ways to see if this holds true for a range of capabilities and morphologies.

The studies presented in this thesis collected data from single encounters. It will be important for longitudinal studies to understand if an intervening robot continues to make these positive differences in the human-human interaction. Longitudinal studies will allow for a more direct assessment of relationship amelioration, and they will allow researchers to understand whether robots designed to support communication in this way fall into disuse or misuse.

The computational model presented in this thesis had moderate positive predictive value for identifying two problematic relationship states. In addition to the implementation and testing of additional model states, there is the opportunity to expand upon the implementation presented here. As natural language processing and computer vision technologies mature to provide a greater contextual understanding of human-human interaction, additional features can be added to enhance our work.

7.3 A Robot to Support Human-Human Communication

A robot to support human-human communication would be invaluable. It's potential to be present at times when human professionals cannot be and ability to recognize issues as they first emerge could transform human-human relationships. This is a complex task, however. The task is one which human professionals still struggle with. This thesis is an important step in making robots capable of supporting human-human communication in hierarchical relationship.

It argues that a robot can positively influence the communication in these types of relationships. Specifically, an intervening robot that makes use of techniques rooted in transformative mediation encourages the use of more explicit emotional expression. In situations when the dyad members are weak, it encourages them to make empowered statements more often than an unengaged robot. When people speak openly about emotion, and they make empowered statements, relationship change for the better (Bush & Folger, 2010a).

Humans do not need a robot to tell them how to communicate. They do not need a robot to tell them how to relate. The technologies suggested here are to support open communication. As George Bernard Shaw said, “The single biggest problem in communication is the illusion that it has taken place.” The technology here does not change the human-human communication; it helps make the human-human communication happen.

APPENDIX A

RATING SCALE FOR EMERGING EMPOWERMENT AND RECOGNITION

This document explicitly defines how the video raters code the participant's (individuals of interest) empowerment and recognition shifts in the audiovisual recordings of their discussions. Coders rate the instances where the participant shows clear evidence of a shift from weakness (e.g. confusion) to strength (e.g. clarity) and from alienation to recognition. There are expected to be more instances of emerging empowerment and recognition in the conditions with an intervening robot than in the conditions with the robot not intervening.

<u>Empowerment Shift</u>

Note: A shift from general and tentative statements to specific and powerful statements shows growing clarity and is indicative of an empowerment shift (e.g. Bush & Folger, 2005; Moen et al., 2001). A shift from being closed off to the other to having an open posture toward (and willingness to hear) the other is an empowerment shift (e.g. Moen et al., 2001).

Coding:

Each time the participant speaks the coder notes if the participant uses language that shows either a lack of empowerment or emerging empowerment. Moen et al. (2001) describes language showing that a lack of empowerment and language that shows emerging empowerment. If the participant is explicit about being uncertain or helpless, like saying "I don't know…" or "There's nothing I can…", this shows a lack of empowerment. If the participant consistently uses tentative language to qualify his/her speech, such as "maybe" or "perhaps", this shows a lack of empowerment. If the participant uses the type of language described in the negativity state (in Appendix B), where he/she uses unreasoned attacks to debase or change the other, then it he/she shows a lack of empowerment. Emerging empowerment is described as speaking in precise/specific terms (Moen et al., 2001). The participant states that this what I believe and what makes me believe it. If the individual brings up a belief or feeling, he/she is explicit in what is making her feel that way. If the participant is critiquing the other, then he/she is specific as to the reason he/she is unhappy with the confederate. For example, Bush and Folger (2005) describe a statement of critique where strength is shown. One party tells the other party directly that she has "painted a picture" of her; the picture of her is "not an accurate picture", and, therefore, she "has a problem with it". This can be contrasted to instances of dyad members using general platitudes or speaking in generalities. If a participant states something to the effect of "this is the way it is…" or "this is just the way I feel…", without any basis or further clarification, this shows confusion and therefore a lack of empowerment. An empowerment shift is a change from a lack of empowerment to emerging empowerment (a change from speaking from a place of confusion to speaking with clarity).

A rating is given to the participant's nonverbal behavior each time he/she is listening to the confederate. This rating is given using a scale developed for a previous study that was approved by a rapport expert. The scale ranged from behavior that was destructive to rapport (1) to optimal for encouraging rapport (4). A change in behavior from behavior that is destructive to a social bond (a 1 or 2) to behavior that is open to a social connection/open to the other's ideas (3 or 4) is an empowerment shift (see Appendix C).

Recognition Shift

Note: Moen et al. (2001) speak about the difficulty of recognizing a "genuine" recognition shift because it depends entirely on the thoughts of the individual who is giving the recognition. Based their experiences, however, Moen et al. (2001) are able to give two different conditions that are frequently true when an individual seems to give recognition and becomes more empathetic toward the other's viewpoint as time moves on. Often, genuine recognition shifts will follow empowerment shifts; shortly after a person shows openness toward the other, they will better understand the other (Moen et al., 2001). At the time of recognition, the individual needs to have the strength (self belief) to be open to ideas that challenge the self's capabilities and competencies. This strength could be gained through an empowerment shift, or the individual may already have the strength. Also, with a genuine recognition shift, the individual does not simply say “yes…but”. That is, the individual truly appreciates the perspective of the other and sees the situation in a novel way. If the individual simply agrees with the other, and moves to back to his/her point, he/she has not had a genuine recognition shift (Moen et al., 2001).

Coding:

The confederate is going to end each statement with a request for recognition; therefore, each utterance of the participant can be rated as showing emerging recognition or refusing recognition according to what is described in Moen et al. (2001). If the participant shows no recognition and then begins to show emerging recognition, this is considered a recognition shift. The segments of text below describe emerging recognition. They are essentially the complement of the cases described in the insensitivity state defined in Appendix B.

In the literature, emerging recognition is shown by explicit acknowledgement of the other’s viewpoint or a positive appreciation of his/her thought processes. For example, Moen et al. (2001) say that an individual might reference the other’s argument, and say, “I’m beginning to understand why you saw it that way”. This type of talk (when participant actually takes the time to discuss the position and rational thinking on the other’s part without immediately following the statement with “but” and returning to one of his/her own arguments) is showing the participant hears the confederate; further, it validates what the confederate is saying because the participant can see feeling/thinking the same way given similar experiences. The participant might also say something about seeing the participant in a new light, e.g. "Now that I see you…" (Moen et al., 2001). Once again, the participant is perspective taking and sees the participant as someone who is not purely antagonistic but has different experiences that have led them to think this way, even if it not the most rational way, from his/her perspective.

Giving serious thought/consideration to the confederate's argument may also show emerging recognition. For example, Moen et al. (2001) describe how an individual might say that a controversial action could be considered "if" it was altered in some way. If the participant uses conditionals and is explicit about trying to find a version of the confederate's viewpoint that is agreeable, this shows an emerging recognition. Similarly, if the participant explicitly says that he/she is comfortable arguing that way, or that argument makes sense to him/her that validates the confederate's viewpoint/feelings, which validates the self.

Finally, if the participant explicitly acknowledges the information the confederate is presenting as novel, then this can be a sign of emerging recognition (Moen et al., 2001). If the participant says something like, "I never realized…" he/she is listening to the confederate, treating him/her as what he/she is presenting is valid, and is willing to let what is being presented change his/her mind.

APPENDIX B

VIDEO CODING INSTRUCTIONS FOR COMPUTATIONAL MODEL STATES OF WEAKNESS AND ALIENATION

State of Weakness:

Negativity

Description: The state of negativity within the model is active when the lower-power individual has strong conscious negative feelings about the self and his/her behavior is demarcated by this negative affect. This is true when the individual displays helplessness and tries to withdraw, but it is also true when the individual gets defensive and separates himself/herself from the other. As explained in Moen et al. (2001), an individual can seem to be sure or even empowered when he/she clearly expresses frustration at the other or seems to be standing his/her ground. This negativity, however, conceals the individual's true vulnerability/weakness, which is communicated in the desire to disengage from the other person or the process. The lower-power individual doesn't feel that he/she has the capacity to resolve the situation. Bush and Folger (2005) describe a piece of an interaction where the tone between parties is not friendly (negative affect is clearly expressed), but the parties use powerful and clear language and show engagement. When there is strength and engagement shown by the dyad, there is a reason not to intervene; it undermines empowerment. There is reason to intervene during the defensive and self-protective behavioral displays or the helpless behavioral displays described in Moen et al. (2001) because the mediator can support empowerment by showing recognition and/or reaffirming the party's capacity to handle the situation.

Coding: When the participant (in a lower-power condition) speaks, this state will either be marked as active or inactive. The negativity state will be active if there is consistent evidence that the individual does not feel capable of continuing to engage in the situation at hand. That is, if there are multiple cues from the categories below (particular attention should be paid to verbal cues).

Verbal Cues

1) The individual is continually repeating the same argument (Moen et al., 2001). Nothing is added to the argument nor is it being stated more clearly. The individual cycles on the same point to avoid further engagement with the other.
2) The individual pushes the other away (makes the other seem impossible to deal with) or tries to force the other to act differently. Retzinger (1991) specifically defines a type of anger or frustration that emerges from shame when powerful feelings of weakness are not discussed directly. The individual challenges the other and tries to lower the status of the other (Retzinger, 1991). This is the frustration described in transformative mediation (e.g. Moen et al., 2001). The individual explicitly blames or criticizes the other for making the situation difficult or being unreasonable (e.g. "You don't understand…!" in Bush and Folger (2010)). The individual demands (could use threats to get) or prescribes the other act in a certain way (e.g. "Get over it!" from Moen et al., 2001). The individual questions the other as if interrogating him/her or assigns them unscrupulous thoughts or motives for his/her position/behavior (e.g. you're doing this simply because "…you don't like it!" in Bush and Folger (2005)). In their example, Bush and Folger (2005) show how an empowered individual can be very critical of the other's behavior. The individual clearly states the specific aspects of the behavior that is undesirable,

what makes it undesirable to her, and makes her feelings, that "she has a problem with it", explicit. The key difference is the explicit acknowledgment of the feelings, the willingness to deal with the situation directly, and the clarity. The individual is not trying to force change or disengage.

3) The individual indicates that he/she is helpless. S/he makes an explicit indication that he/she does not wish to continue in the interaction or shuts down. For example, Bush and Folger (2010) gives the example of "I've had enough of this"! Moen et al. (2001) gives the example "There's nothing I can do…".

Paralinguistic Cues

1) Retzinger (1991) lists paralinguistic cues for shame. Each cue is an indicator of the negativity. The individual speaks very softly, frequently pauses or interrupts himself/herself. During pauses the individual may use common fillers, like "um", "like", etc. The individual talks very fast or has irregular speech rhythm. Finally, one might also laugh at certain words or laugh during obviously tense moments.
2) Retzinger (1991) also lists paralinguistic cues for anger. The individual speaks loudly, putting heavy stress on certain words. The individual may alternatively speak with a whiny quality to the voice or speak in a singsong way.

Nonverbal Cues

1) Retzinger (1991) lists nonverbal cues of shame. Look for the lower power individual to avert his/her gaze, keep eyes away from the other,

or turn away from the other. S/he could may have his/her hands up around and covering parts of the face.

2) Retzinger (1991) lists nonverbal cues of anger. Look for the individual to have narrowed eyelids and/or bulging eyes as he/she glares at the other. Additionally or alternatively, may lean to toward the other (in a challenging way). Look for the lips to be tightly pressed together.

<u>State of Alienation:</u>

Insensitivity

Description: The insensitivity state is active if the higher-power individual shows a lack of positive affect for the lower-power dyad member. This is true when the individual is actively antagonistic toward the other; however, it is also true when the higher-power dyad member demonstrates no interest in the other when the other makes a request for recognition (either case shows a "refusal" of recognition) (Moen et al., 2001). When the participant is listening to the confederate speak, he/she can nonverbally demonstrate that he/she is receptive to what they are saying, or they can communicate a disinterest or distaste for what the confederate is saying. These types of nonverbal communication pose problems for the pair developing and maintaining a social bond (Tickle-Degnen, 2006).

Coding: This state can be active or inactive when the participant is listening to the confederate or when the participant is responding to a request for recognition from the confederate (when the participant is in the higher-power role). The state is marked as active or inactive for each of these situations.

When the participant is speaking, the insensitivity state is active when the participant does not show genuine recognition as described in Appendix A. That is, the state is active when the participant refuses recognition after the confederate has made a request for recognition. Moen et al. (2001) describes how to spot the refusal of recognition. If any of these three cases are true, the insensitivity state is active when the participant is talking.

1) The participant simply pays “lip service” to the position of the confederate. The participant may say “sure” or “that makes sense” in response to the confederate’s point; however, the participant almost immediately says “but” and/or moves onto his/her next point. Essentially, this case encompasses any situation where the confederate has made a request for recognition, and the participant has reacted with disinterest.
2) The participant could actively antagonize the other and dismiss his/her ideas. For example, Moen et al. (2001) describes making accusations, using sarcasm, and name-calling as tactics of dismissing the other’s point of view. The text gives the example of using “that’s original” sarcastically. Sarcastic comments and name-calling calls into question the individual’s competency; it suggests listening to the ideas is a waste of time. This weakens the other individual. Moen et al. (2001) additionally give examples of an individual “minimalizing or trivializing” the ideas of the other party or assuming the “worst motives” of the other party. The individual might say something like “she/he only wants to…” or “as usual…” (Moen et al., 2001). The participant would be stereotyping the confederate and showing an unwillingness to hear his/her thoughts and feelings because he/she is already aware of what’s to come, and it is unmoving or unpersuasive. The participant is weakened by a clear belief in his/her lack of competency and a lack of care.
3) The participant could also be explicit in his/her unwillingness to hear the perspective put forth by the confederate (Moen et al., 2001). The

> unwillingness to even consider the other person's point of view tells that individual that his/her ideas are unworthy of consideration, which brings into question his/her value and further weakens him/her.

The participant's nonverbal behavior will be rated when he/she is listening to the confederate. The participant's behavior will be rated using the expert approved scale developed from the rapport literature (Tickle-Degnen, 2006). This scale can be found in Appendix C. If the participant demonstrates closed off nonverbal behavior while listening (showing self-consumption or disinterest toward the other), which is indicated by a 1 or 2 on the scale, then the insensitivity state is active. If the participant is actively engaged and responsive to the other, a 3 or 4 on the scale, then the state is not active.

APPENDIX C

RATING SCALE FOR NONVERBAL BEHAVIOR

This document gives explicit rating instructions to coders who will rate the nonverbal behavior of the participant (an individual of interest). The participant's behavior will be rated each time the confederate is making an argument when the participant is in the higher-power role. If the participant is showing rapport damaging behavior, then the insensitivity state is active. If the participant's behaviors are associated with being open to the other (behaviors associated with optimal rapport), then the insensitivity state is not active. The insensitivity state is explicitly defined in the Appendix B.

Tickle-Degnen (2006) gave dimensions of nonverbal behavior that influence rapport. This scale, influenced by Tickle-Degnen (2006) and approved by Tickle-Degnen, was used to code the participant's behavior each time the confederate is making an argument. Video coders will give a rating of a rating of 1, 2, 3, or 4. They will indicate whether the behavior that damaged the rapport in the interaction was "over the top" or due to participant being withdrawn by using (a or b). A rating of 4 is the "optimal experience" for rapport.

Rating of Participant's Nonverbal Behavior – "Optimal Experience" for Rapport

1a = The person is very clearly not "in sync" with the other. There may be almost no time spent looking in the direction of the other, and he/she may be physically withdrawn/oriented away from the other. Nonverbal responses (i.e. head nods, smiles or recognizable emotional facial responses) to what the other is saying are absent or inappropriate.

1b = Alternatively, the person may appear "on edge" to the point of making the other uncomfortable (i.e. lots of fidgeting and somewhat unnecessary movement). S/he may be leaning toward the other, giving him/her an "unnatural" attention (seemingly examining him/her). The individual's nonverbal responses are "over the top" (i.e. emphatic gesturing or overt, recognizable emotional response that is excessive) and may seem irregular.

2a = The person may make some acknowledgement of the other, but it is clear his focus is elsewhere. S/he may make some gestures such as nodding at the other as he/she speaks, but he/she is not showing consistent attention or showing understanding through nonverbal cues. The individual is often looking away from the other, and his/her orientation is not directly toward the other.

2b = Alternatively, the person appears nervous; he/she may fidget slightly or sit rigidly. The individual's eyes don't really leave the other. S/he makes consistent acknowledgement that he/she is listening by nodding head repeatedly. His nonverbal emotional response may be excessive relative to what the other person has said.

3a = The person may glance away from the other, but he is regularly looking toward the other. S/he is oriented toward the other and has a relaxed posture. When the other is saying something particularly meaningful/personal, the individual is looking at the other, nodding for acknowledgement, and showing an appropriate emotion.

3b = Alternatively, the individual may be fidgeting slightly. S/he is oriented toward the other and his/her posture is more relaxed (not totally rigid). He may nod his head or use other indicators to show he is listening in excess. His emotional response to personal disclosures is appropriate and not excessive.

4 = The person is relaxed. S/he generally maintains eye contact with the other (without staring or glaring) and is oriented toward the other. S/he is relaxed; his posture is not rigid, and he does not lean toward the other. S/he does not fidget or make excessive, unnecessary movements. S/he does not nod excessively but to indicate agreement and/or understanding to the other. The individual's emotional responses are natural and reasonable given what the participant has said.

APPENDIX D

IRB-APPROVED CONSENT FORMS, SELF-REPORT MEASURES, AND FLYER

This appendix includes the consent and debriefing forms for the studies described in chapters 5 and 6, the measures given to participants, and the study's advertisement.

GEORGIA INSTITUTE OF TECHNOLOGY
CONSENT TO BE A RESEARCH PARTICIPANT

Project Title: Human-Robot Interaction Study: Supporting Human-Human Interactions

Investigators: Arkin, R.C., Ph.D. and Pettinati, M. J.

Protocol and Consent Title: Using robots to support human-human relationships consent form for adult participation

You are being asked to be a volunteer in a research study.

Key Information:
You will be asked to form an argument with another participant in this study. This argument is whether or not concealed carry of a gun on the Georgia Tech campus should be allowed. Your discussion will be recorded. Your compensation will depend on the strength of this argument. Forming the argument may be stressful at points because of the topic's controversial nature. You may not want to be in this study due to this discomfort. You may want to participate because you have the chance to talk about a topic you have strong feelings about. Also, it contributes to our understanding of how a robot can help people relate.

Study Purpose:
There are times when people disagree with one another. In these moments, a robot may be able to help people get along. This study is about how the robot can recognize what's happening between two people. Also, how a robot might help people relate.

Exclusion/Inclusion Criteria:
Participants in this study must be over the age of 18 and under the age of 70.
Participants must currently be enrolled as students at the Georgia Institute of Technology.
Participants must have a strong opinion on the statement: "I feel less safe on the Georgia Tech campus with it being legal for permit holders to conceal carry a gun, and it should not be allowed".

Procedures:
During the study, you may be led to believe some things that are not true. When the study is over, we will tell you everything. At that time you can decide whether to let us use your information. You can also stop the study at any time for any reason. You have the right to require that your information be destroyed and not be used in the study.

Page 1 of 5
Adult Consent Template March 2017

APPROVED *Consent Form Approved by Georgia Tech IRB: February 15, 2019 - February 14, 2020*

If you chose to participate in this study, your part will only require this single one-hour visit to the Technology Square Research Building. You will be randomly assigned to one of four different groups.

You will begin the study by completing surveys that collect demographic information. After completing the surveys, you and the other participant will be led into TSRB 238A (Dr. Ronald C. Arkin's office). You will be asked to wear an Empatica E4 Wristband. This device measures things like skin temperature and heart rate. You will also be asked to wear a microphone. With your permission, we will collect and analyze the information from these devices. You will then have three minutes to relax before the experiment progresses.

After the relaxation period, you and the other research participant will be given information about the study. You will be asked to discuss the issue of concealed carry on the GT campus. You will be asked to formulate an argument for one side of the issue. Either you or the other research participant will be assigned to present the argument at the interaction's conclusion. Your compensation depends on the presented argument. You and the other participant will be given 15 minutes to create the argument. If you brought notes for the discussion, you may refer to these throughout the discussion. You may also take notes during the discussion. The discussion will be organized. You will have one to two minutes to make a point and respond to the other party without interruption. The other participant will then have one to two minutes to respond you and make his/her next point without interruption. This pattern will continue until the end of the fifteen minutes. This will help to ensure a fair and equitable discussion.

With your permission, we will record your interaction with the other participant. The recording devices will be recording when you enter the office.

After the discussion, you or the other participant (whomever was assigned) will make an argument on behalf of both of you. The presenter can refer to his/her notes while presenting. The person not presenting should remain silent during the presentation of the argument. After you or the other participant makes this argument, you will complete surveys. The surveys relate to your impression of the discussion, your partner, and the robot.

After you have completed the post-interview surveys, both of you will be debriefed. Any questions you have about the study or our project will be answered. You will also have the opportunity to confirm that we are able to use your data.

Finally, we ask that you do not discuss this study with anyone for the next six months. Also, please do not post about the study on social media. It is important that future subjects do not know the details of the study.

Risks or Discomforts:
You will be asked to discuss a controversial topic. When discussing this topic, you may experience some emotional discomfort. However, this emotional discomfort should be no greater than participating in a debate in an ethics class. Remember, you are also able to stop the study at any time without negative repercussions.

Benefits:
You are not likely to benefit in any way from joining this study. We hope that what we learn from this study will help to develop robotic systems that support human relationships. These robotic systems will help the two people have more productive and positive (dignified) interactions.

Compensation to You:
You will be compensated $5-$15 cash for your time and effort.
U.S. Tax Law requires that a 1099-misc be issued if U.S. tax residents receive $600 or more per calendar year. If non-U.S. tax residents receive more than $75, mandatory 30% withholding is required. Your address and Tax I.D. may be collected for compensation purposes only. This information will be shared only with the Georgia Tech department that issues compensation, if any, for your participation.

Use of Photographs, Audio, or Video Recordings:
With your permission, we will record the discussion between you and the other participant. The recordings created will be audio and visual recordings. You will be identifiable in the recording. We may want to use the recordings for public presentations related to our research. The attached MODEL RELEASE FORM outlines several possible uses. It asks for your written consent to use these items in each way. We will not use any identifiable information about you in any future presentation or publication without your consent. If you do not sign the attached MODEL RELEASE FORM, then only study staff will view and have access to the recordings. The recording will be kept under lock and key.

Confidentiality:
The data collected about you will be kept private to the extent required by law. To protect your privacy, your records will be kept under a code number rather than by name. These records, including the recordings of your discussion, will be kept in a locked file cabinet. The recordings will be stored on an external hard drive. Only study personnel will have access to this cabinet.

There are four video recording devices that will be used during the interview. Three recording devices will be setup in the office. One recording will be made using the office's desktop's webcam. These will be running when you enter the office.

The recordings will be moved to two external hard drives. This will be done at the end of the day or when the camcorders' memory cards are full (whichever occurs first). An experimenter external to the interview will monitor the interview using the desktop's webcam. This computer will be connected to the internet for the extent of the interview. However, this is a College of Computing password protected machine. This makes the risk for unauthorized access to this machine minimal. The experimenter monitoring the interview will be doing so via an Apple FaceTime call. The data transferred during an Apple FaceTime call is encrypted "end-to-end". FaceTime content is also not stored to any servers.

If you have given us permission to use your recordings, we may use them in academic presentations. Otherwise, these recordings will not be presented publicly; they will only be available to the experimenters. We will not report your name or anything that might point to you when publishing results.

To make sure that this research is being carried out in the proper way, the Georgia Institute of Technology IRB may review study records. The Office of Human Research Protections may also look over study records during required reviews.

Future Data Use:
All of your data will always be kept under lock and key. We will destroy your identifiable data after our analysis is complete if you have not given us permission to use your recordings in academic settings (e.g. conferences). If you have given us permission to use your recordings, we will erase your data within ten years of the study's completion. A member of the Mobile Robot Lab may use the data that does not identify you in future research studies without additional consent. We will not distribute your data to other researchers.

Costs to You:
There are no costs to you, other than your time, for being in this study.

Conflict of Interest:
None

Questions about the Study:
If you have any questions about the study, you may contact Principal Investigator, Ronald C. Arkin, Ph.D., at arkin@cc.gatech.edu.

In Case of Injury/Harm:
If you are injured as a result of being in this study, please contact Principal Investigator, Ronald C. Arkin, Ph.D., at arkin@cc.gatech.edu. Neither the Principal Investigator nor Georgia Institute of Technology has made provision for payment of costs associated with any injury resulting from participation in this study.

Questions about Your Rights as a Research Participant:

- Your participation in this study is voluntary. You do not have to be in this study if you don't want to be.
- You have the right to change your mind and leave the study at any time without giving any reason and without penalty.
- Any new information that may make you change your mind about being in this study will be given to you.
- You will be given a copy of this consent form to keep.
- You do not waive any of your legal rights by signing this consent form.

If you have any questions about your rights as a research participant, you may contact

Ms. Melanie Clark, Georgia Institute of Technology
Office of Research Integrity Assurance, at (404) 894-6942.

If you sign below, it means that you have read (or have had read to you) the information given in this consent form, and you would like to be a volunteer in this study.

____

Participant Name (printed)

____

Participant Signature Date

____

Signature of Person Obtaining Consent Date

This was the optional video release form given to participants. It was given to them immediately following the consent form; they were explicitly told it was optional.

Georgia Institute of Technology

Release, Waiver of Liability, and Covenant Not to Sue (Model Release)

For good and valuable consideration, the undersigned hereby grants the GEORGIA INSTITUTE OF TECHNOLOGY ("GIT") and the GEORGIA TECH RESEARCH CORPORATION ("GTRC") the absolute and irrevocable right and permission, in respect to the photographs, video tapes, motion pictures, recordings, or any other media (hereinafter collectively known as "Images") that GIT/GTRC has taken of me or my property, or minors in my care, or in which I may be included with others, to copyright the same, in GIT/GTRC's own name or otherwise, to use, re-use, publish, re-publish, and allow others to use, re-use, post, display, publish, re-publish, reproduce the same in whole or part, individually or in conjunction with other images, and in conjunction with any printed or electronic matter, in printed form, microfilm, electronic databases, CD-Rom, websites, online digital archives, or in any and all media now or hereafter known, and for any legitimate purpose whatsoever, and to use my name in connection therewith if GIT/GTRC so chooses. I hereby waive any right to inspect or approve the Images or any finished version incorporating the same.

The undersigned does hereby release and forever discharge GIT, GTRC, and the Board of Regents of the University System of Georgia, their members individually, and their officers, agents, and employees of any kind from all claims, demands, rights, and causes of action of whatever kind or nature, arising from and by reason of any and all known and unknown, foreseen, and unforeseen injuries, damages, and the consequences thereof resulting from the use of the Images, including without limitation any and all claims for libel or invasion of privacy.

I understand that the acceptance of this release and waiver of liability by the Board of Regents of the University System of Georgia shall not constitute a waiver, in whole or in part, of sovereign immunity by said Board, its members, officers, agents, and employees.

This authorization and release shall also inure to the benefit of the heirs, legal representatives, licensees, and assigns of GIT, GTRC, and the Board of Regents of the University System of Georgia. I hereby certify that I am suffering under no legal disabilities and that I have read the above carefully before signing. This release shall be binding upon me and my heirs, legal representatives, and assigns.

By signing below, I agree to the terms stated above and hereby certify that I am 18 years of age or older.

____
Signature/Date

____
Print name

Model Release Form
GT Legal Rev. 1/19/11

This is the debriefing form completed by participants at the conclusion of the study to verify that they understood the purpose of the study and why it had been designed the way it was. The IRB requires consent to be given at the study's conclusion when deception is incorporated into the study. There were four versions of this form. Each version was identical aside from identifying the participant as the lower or higher-power relationship member and identifying the robot as idle (unengaged) or intervening.

Study Debriefing Form

Study Purpose

People may discuss their relationships in terms of roles. A patient could discuss his or her "caregiver". A student could refer to his or her "advisor". These roles imply who has more power when making decisions. For example, a caregiver may have more say than the patient about the plan to treat that patient. The patient has to trust the caregiver will provide the best treatment he or she can. Relationships with power differences can have problems arise. The high-power member in the relationship may not listen to the low-power member. The low-power relationship member may just agree with the higher-power member. These unhealthy ways of working together mean the lower-power member's needs are not met.

We propose that a robot can help to monitor these relationships and help both parties feel heard. This study is about understanding how a robot can best provide this help. We want to see if robots can identify specific problems. We also want to see if this robot can help support better relationships. A robot can amplify the emotion of people in the discussion at certain points. A robot may help a person see how he or she presents himself or herself. A robot might also help one person see how he or she makes the other person in the relationship feel. When the robot acts, it may cause a relationship participant to change his or her thinking and/or behavior.

Study Condition

In this experiment, you each were assigned roles. You had the higher-power role. You made the final argument. You were told the compensation for both of you depended upon this argument. Your partner had low power. S/he could just give advice about the argument. You should know that no matter what your argument was we would have compensated you $15. We told you that your argument determined your payment to create a power difference. Half the participants are given the higher-power role. Half are given the low-power role.

You were in a group with an intervening robot. The robot remained idle until it believed there was a problem in the relationship. It tried to support you to feel stronger in moments when strength was needed. It encouraged you to express yourself. It may have tried to show you how you were expressing yourself. This is so you could be confident in your self-expression. If you did not respond to the thoughts and/or feelings of your partner, it tried to encourage you to see the other's perspective. Half of the participants have this robot. The other study participants have an idle robot that never intervenes.

Additional Study Deception

You may have noticed some tension in the discussion. Your partner caused this tension by beginning the discussion unwilling to hear your ideas and being combative. S/he was critical and inattentive. Your partner did this because he/she was a study confederate. The way he/she acted was scripted and rehearsed. S/he was included in the study to make you feel unsure and frustrated at the beginning of the discussion. We needed you to experience this discomfort to understand if a robot can recognize problems and help fix these problems in relationships with power differences. After the beginning of the discussion, your partner mirrored your response. If you didn't recognize his/her viewpoint, he/she didn't recognize your viewpoint. If you did recognize his/her thinking, then he/she recognized your point of view.

Because you were deceived about your partner's relationship to the experiment and the compensation guidelines, you now have the right to refuse to allow your data to be used and to ask that they be destroyed immediately. If you do so, there is no penalty.

___ I give permission for my data to be used in the analysis for this experiment.

___ I do NOT give my permission for my data to be used in the analysis for this experiment. Please withdraw them from the study and destroy them immediately.

Participant Name (Printed) ______________________________

Participant Signature ______________________________

Date ______________________________

This is the demographics and personality survey that was administered to participants before the discussion took place (immediately after the consent and video release forms were completed by participants). While this was administered at desks outside the office, the experimenter finished setting up the office (e.g. turned on recording devices).

Participant # ______

Demographics and Personality

1. What is your gender?
   ☐ Female ☐ Male

2. What is your age? ___________

3. What is the highest level of education you've achieved?
   ☐ High School ☐ Bachelor's ☐ Master's ☐ Ph.D. ☐ Other ___________

4. How well do you know the other participant?
   ☐ I don't know the other participant ☐ Casual Acquaintance ☐ We know each other well

5. To what degree do you keep up with the news, politics, and societal issues?
   ☐ I don't keep too close an eye on current events.
   ☐ I watch or read the news once in a while.
   ☐ I watch or read from a variety of news sources regularly.
   ☐ I watch or read from a variety of news sources regularly and actively engage others in discussion about various issues.

6. Have you ever interacted with robots? Please check all that apply.
   ☐ Never
   ☐ Very limited interaction
   ☐ Interaction experience with military robots
   ☐ Interaction experience with industrial robots
   ☐ Interaction experience with entertainment/educational robots
   ☐ Interaction experience with humanoid robots
   ☐ Other – please specify ____

7. Do you have expertise in robotics and/or artificial intelligence?
   ☐ I have only been exposed to artificially intelligent agents in popular culture.
   ☐ I am interested in artificial intelligence and robotics and have done some casual independent study.
   ☐ I have taken college-level courses or undertaken significant personal projects related to AI/robotics.
   ☐ I am a robotics/AI researcher.

8. Have you participated in another experiment with this robot in this lab?
   ☐ Yes ☐ No

PLEASE TURN PAPER OVER

Indicate how much you agree/disagree with the following statements by circling a rating for each statement based on the scale shown below.

	Strongly Disagree 1	Disagree 2	Slightly Disagree 3	Neutral 4	Slightly Agree 5	Agree 6	Strongly Agree 7
9) I like to lead when having a group discussion.	1	2	3	4	5	6	7
10) I go along with other people's ideas when deciding what to do.	1	2	3	4	5	6	7
11) I am comfortable expressing my opinions on a contentious topic.	1	2	3	4	5	6	7
12) I am usually confident in how I express my ideas and myself.	1	2	3	4	5	6	7
13) When I disagree with a superior, I often feel obligated to put their needs and wants above my own.	1	2	3	4	5	6	7
14) When I openly disagree with someone who has more control than I do in a situation, I often feel like my ideas aren't seriously considered.	1	2	3	4	5	6	7
15) When I have strong feelings about a topic or decision, it can be difficult to listen to and appreciate alternative viewpoints.	1	2	3	4	5	6	7
16) Sometimes two people who disagree need a neutral third-party to help them listen to one another and come to mutually satisfactory resolution.	1	2	3	4	5	6	7
17) If I were having a dispute with someone, a robot's presence would make me more uncomfortable than a third person.	1	2	3	4	5	6	7
18) A robot could never accurately understand and reflect how I'm feeling or how I'm behaving.	1	2	3	4	5	6	7
19) If a robot were to try to reflect my feelings, I would feel like I was being judged.	1	2	3	4	5	6	7
20) A robot couldn't do anything that would help me open up to or better identify with another person.	1	2	3	4	5	6	7

This is the self-report form given to participants to gather their impressions of the robotic agent. They completed this form after the discussion. This was completed along with the empowerment and recognition items (below) as well as the previously validated measures (used with permission), Jehn's (1995; 2001) intragroup conflict scales and Burgoon and Hale's (1984; 1987) Relationship Communication Scales.

Participant #: ________

Please respond by circling 1, 2, 3, 4, or 5 to indicate how strongly you agree with the following statements.

	1 Strongly Disagree	2 Disagree	3 Neutral	4 Agree	5 Strongly Agree
1. The robot was disruptive.	1	2	3	4	5
2. The robot clearly didn't understand how I felt.	1	2	3	4	5
3. The robot was on my side.	1	2	3	4	5

Any Additional Comments or Thoughts About the Robot:

This is the survey containing empowerment and recognition items that were inspired by questions asked by practicing transformative mediators to assess how parties felt the mediation session had progressed (Folger, 2010). This was completed after the discussion along with the robot impression measures (above) and the previously validated measures (used with permission), Jehn's (1995; 2001) intragroup conflict scales and Burgoon and Hale's (1984; 1987) Relationship Communication Scales.

Participant #: _______

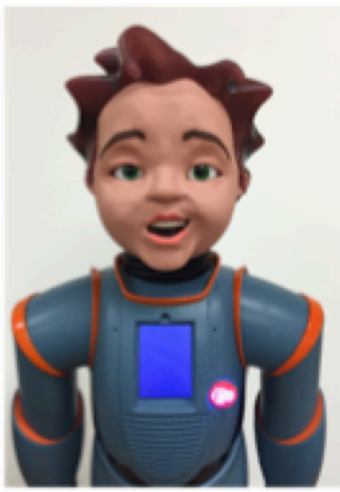
Please respond by circling 1, 2, 3, 4, or 5 to indicate how strongly you agree with the following statements.

	1 Strongly Disagree	2 Disagree	3 Neutral	4 Agree	5 Strongly Agree
1. I had the opportunity to express myself.	1	2	3	4	5
2. I had the opportunity to make my points and ask the questions I had of the other participant.	1	2	3	4	5
3. This process helped me become clearer about my thoughts and my feelings on the issue.	1	2	3	4	5
4. I gained a better understanding of the other person during the process.	1	2	3	4	5
5. I was able to see the other person's perspectives, views or opinions.	1	2	3	4	5
6. The other person was able to see my perspective, views or opinions.	1	2	3	4	5

This was the advertisement that was hung around campus to advertise the study to potential participants. The text included on this advertisement was the same as the text in emails sent to mailing lists used to recruit participants.

Georgia Tech | College of Computing
Defining the New Face of Computing

OPINIONATED STUDY PARTICIPANTS WANTED!!!



The Mobile Robot Laboratory in the Tech Square Research Building (TSRB) is conducting a human-human-robot interaction study. Do you have a strong opinion on individuals being able to conceal carry a gun on the Georgia Tech campus? Would you be willing to share your thoughts on this issue? During this study, you will be asked to construct an argument with one other participant on this issue. You will also get to work with one of our humanoid robots, Milo! This study requires no more than 60 minutes of your time, and we'll give you $5-$15 cash for taking the time to help us out. In order to qualify for this study, you must be at least 18 years of age and under 70 years of age, and you must currently be a student who is enrolled at Georgia Tech. It is not necessary to have any technical background to participate, and people with no robotic experience are strongly encouraged to participate (though all backgrounds are welcome).

Thank you for your attention! We look forward to hearing from you!

If you are interested in participating, please contact Michael Pettinati at **mpettinati3@gatech.edu**.

GA Tech Mobile Robot Lab Study
mpettinati3@gatech.edu

GA Tech Mobile Robot Lab Study
mpettinati3@gatech.edu

GA Tech Mobile Robot Lab Study
mpettinati3@gatech.edu

GA Tech Mobile Robot Lab Study
mpettinati3@gatech.edu

GA Tech Mobile Robot Lab Study
mpettinati3@gatech.edu

GA Tech Mobile Robot Lab Study
mpettinati3@gatech.edu

GA Tech Mobile Robot Lab Study
mpettinati3@gatech.edu

GA Tech Mobile Robot Lab Study
mpettinati3@gatech.edu

GA Tech Mobile Robot Lab Study
mpettinati3@gatech.edu

GA Tech Mobile Robot Lab Study
mpettinati3@gatech.edu

GA Tech Mobile Robot Lab Study
mpettinati3@gatech.edu

GA Tech Mobile Robot Lab Study
mpettinati3@gatech.edu

GA Tech Mobile Robot Lab Study
mpettinati3@gatech.edu

APPENDIX E

OUTCOME MEASURE GENERATION, EXPERIMENT SENSORS AND EXPERIMENTER BIAS

This appendix has three distinct parts. The first part discusses the generation of and purpose for the outcome measures, the values that were used to assess the computational architecture and research questions described in the thesis. The second section describes the sensors that were used to collect all of the experiment data and how that data was used. Finally, the fact that an experimenter was directly involved in the data generation process introduces experimenter bias. The final section this appendix discusses this issue, measures that were taken to mitigate it as well as evidence that these data are reliable.

Outcome Measure Generation

<u>Objective Measurements</u>

- The participant's total speaking time (seconds)

  Generation: Experiment participants and the confederate had structured discussions about concealed carry on GT. As described in Table 5.11, they had time to speak uninterrupted to make their point and respond to the confederate's points before yielding the floor to the confederate who then had time to speak uninterrupted. This structured exchange took place for the entire discussion. Using the audiovisual recordings, the experimenter computed the number of seconds from when the participant began to speak (took the floor) to the second he/she stopped speaking and allowed the confederate to make a point (yielded the floor). The sum of these time periods is the total speaking time for the participant.

  Purpose: An increased speaking times shows an increased engagement in the conversation and increased engagement in the conversation can show greater participant empowerment (Bush & Folger, 2010a).

- The participant's average speaking time (seconds)

  Generation: This is the participant's total speaking time divided by the number of speaking turns that the participant had.

  Purpose: The participant's total speaking time could contain many short speaking sections or a few long speaking sections. Very long or very short speaking sections could indicate less engagement as the participant could be treating the discussion as competing monologues or indifferent toward what has been said (Bush & Folger, 2010a).

- The length of the final argument by a higher-power participant (seconds)

  Generation: As described in chapter five, higher-power participants made a final argument on which the compensation for both themselves and for the confederate was allegedly based. The experimenter computed the number of seconds from when they started speaking to the second the stopped speaking when making this final argument.

  Purpose: Participants who engaged in the conversation, who heard and made persuasive conversations on both sides of the issue may make a longer final argument than participants who were less engaged.

- The count of unique arguments presented by the participant during the discussion

  Generation: The guidelines provided to the study confederate, which are available online[16], explicitly enumerated arguments on both sides of the conceal carry debate drawn from a variety of sources. The experimenter counted those arguments that were presented by a particular participant.

  Purpose: Repeated arguments and getting stuck on a single argument can be a sign of weakness in the participants (Bush & Folger, 2010a). Stronger participants may present more unique arguments during the discussion.

- The count of unique argument presented by the higher-power participant during the final argument

  Generation: The experimenter used audiovisual recordings to count the arguments presented during the final argument by higher-power participants (from the same list as the above count).

---

[16] https://www.cc.gatech.edu/ai/robot-lab/nri_thesis/thesis_supplemental/Guidelines.docx

Purpose: If the participant is strong during the discussion, he/she is likely to be able to recognize more ideas that were presented from the study confederate (Bush & Folger, 2010a). This might mean presenting more arguments during the final argument.

- The number of times the participant mentioned his/her feelings (how he/she was presently feeling)

  Generation: The experimenter used audiovisual recordings to count the number of instances in which the participant explicitly affirmed his/her emotional state or verbally stated, "I feel…", "I am…" followed by an emotion (e.g. uncertain, uncomfortable).

  Purpose: Acknowledgement of emotion is an important part of overcoming weakness and moving into strength (Bush & Folger, 2010a; Retzinger, 1991). There may be more opportunities for empowerment or more empowered participants if participants mention emotion more.

- The experimenter counted the number of times that the robot was explicitly

  Generation: The experimenter counted each time the participant made an explicit verbal reference to the robot using the audiovisual recordings.

  Purpose: If participants are mentioning the robot frequently during the conversation, the robot may have been distracting or its purpose may not have been clear to the participants. The robot was meant to be peripheral and not actively involved in the conversation.

- The participant's number gazes toward the robot

  Generation: There were visual recordings of the discussion that provided a full faced view of experiment participants. The robot was positioned at the periphery of the discussion such that participants had to turn their heads to gaze at it. The experimenter counted the number of times that the participants turned their heads to gaze at the robot.

  Purpose: If participants were spending a lot of time looking at the robot during the discussion, the robot may have been distracting or its purpose may not have been clear to the participants.

Physiological Measurements

- blood volume pulse
- skin temperature
- skin conductance

  Generation: These were collected directly from sensors built into the Empatica E4 Wristband's[17].

  Purpose: Physiological reactance has been shown to be correlated with the onset of emotional stress (Bradford & Lang, 2000). So, for example, spiking skin conductance measurements may have been present in participants who were experiencing disempowerment.

Self-Report Measures

- Perceived Emotional Conflict
- Perceived Task-Related Conflict

  Generation: These data were directly collected from pen and pencil surveys completed by participants immediately following the discussion. These surveys were used with permission from author and publisher (Jehn, 1995; Jehn & Mannix, 2001).

  Purpose: These surveys have been employed in many different contexts that look at group and pair functioning. It was important to have an understanding of the participants' experiences of the conflict and whether that matched with what would be anticipated given existing literature as well as other measures gathered as part of the study.

- Partner's perceived involvement and affection shown during the conversation
- Partner's perceived receptivity to ideas and trustworthiness
- Partner's perceived depth/similarity/inclusion
- Partner's perceived dominance
- Partner's composure
- Partner's formality
- Partner's task vs. social orientation

  Generation: These data were directly collected from pen and pencil surveys completed by participants immediately following the discussion. These surveys were used with permission from author and publisher (Burgoon & Hale, 1984 & 1987).

[17] https://www.empatica.com/en-eu/research/e4/

Purpose: It was important to have feedback from the participants regarding the experiment confederate. Empowered participants might be more open and honest regarding their experience with the confederate (Bush & Folger, 2010a). Robotic interventions can make participants more aware of the problematic dynamics of the interaction with the confederate. Further, the power dynamic in the relationship could make higher- and lower-power dyad members perceive their experiences with the confederate differently.

- Participant's perceived level of empowerment and recognition

  Generation: These are taken directly from pen and pencil survey questions that are asked after the discussion. These questions are based on those asked by human mediators following sessions (Folger, 2010). The questions appear in Appendix D.

  Purpose: These relate to how participants' experienced the conflict and act to corroborate the video-rating data.

- The robot was a distraction to the participants

  Generation: This was taken directly from a single pencil and paper survey question asked after the discussion (appears in Appendix D).

  Purpose: It was important to understand if the participant saw the robot as distracting. Higher-power participants may have seen the robot as more of an adversary and paid more attention to the agent. Interventions may have been seen as distractions by participants.

- Participants believed the robot understood them

  Generation: This was taken directly from a single pencil and paper survey question asked after the discussion (appears in Appendix D).

  Purpose: If the robot is viewed as not understanding the dyad members, then the interventions or the robot may fall into disuse, particularly over time (e.g. DiMicco et al., 2004).

- The robot was perceived to be on the participant's side

  Generation: This was taken directly from a single pencil and paper survey question asked after the discussion (appears in Appendix D).

  Purpose: It is important for the robot to not be on the participant's side. The robot is meant to be perceived as neutral (Bush & Folger, 2010a). The mere presence of the robot could make higher-power participants feel they are being monitored. The interventions could make the robot seem biased to one side or the other. This was used to ensure higher- and lower-power participants viewed the robot similarly and that interventions did not bias the robot. Interventions were designed to be neutral.

Video-Rating Measures

- Video Segments in which the participant was speaking or listening were labeled as showing a lack of empowerment (or emerging empowerment)

  Generation: The exact way in which these segments were labelled by the experimenter and the video coder is explicated in Appendix A of the thesis.

  Purpose: If the proportion of video segments that show weakness, is higher in higher- or lower-power participants, then the dynamics of conflict described in the literature have been altered. The conflict has become asymmetrical. If the proportion video segments that show weakness are higher in the unengaged robot condition, described in chapter five, then the interventions may have successfully supported the relationship moving out of destructive conflict. If the proportion of video segments that show strength following weakness are higher in the intervening robot condition, then the interventions may have supported empowerment shifts (Bush & Folger, 2010a).

- Video Segments in which the participant was speaking were labeled as showing the state of negativity or not (the state was active or inactive)

  Generation: The exact way in which these segments were labelled by the experimenter and the video coder is explicated in Appendices B and C.

  Purpose: The state of negativity is one problematic relationship state identified in the computational model in chapter three. There have not been any computational systems validated to identify this state. There are no datasets with ground truth labels of this state that have been collected along with sensor data. The definition of this state includes identifying the cooccurrence of verbal, paralinguistic and nonverbal cues (as described in Chapter 4 and Appendix B). The hand labeled states served as a ground truth from the which the computational parameters could be learned and tested as described in chapter five section eight.

- Video Segments in which the participant was speaking were labeled as showing a lack of recognition (or genuine recognition)

  Generation: The exact way in which these segments were labelled by the experimenter and the video coder is explicated in Appendix A of the thesis.

  Purpose: If the proportion of video segments that show alienation is different for higher- and lower- power dyad members, then the conflict is not symmetric (as it is described in the literature). If the proportion of video segments that show recognition is higher in the intervening conditions, then the interventions may have been successful in helping to shift the relationship out of destructive conflict. If a higher proportion of video segments in the intervening condition show recognition (after the participant has refused recognition), then the interventions may have helped to support recognition shifts (Bush & Folger, 2010a).

- Video Segments in which the participant was speaking or listening were labeled as showing the state of insensitivity or not (the state was active or inactive)

  Generation: The exact way in which these segments were labelled by the experimenter and the video coder is explicated in Appendices B and C.

  Purpose: The state of insensitivity is one problematic relationship state identified in the computational model in chapter three. There have not been any computational systems validated to identify this state. There are no datasets with ground truth labels of this state that have been collected along with sensor data. The definition of this state includes identifying the cooccurrence of verbal, paralinguistic and nonverbal cues (as described in Chapter 4 and Appendix B). The hand labeled states served as a ground truth from the which the computational parameters could be learned and tested as described in chapter five section eight.

Experiment's Sensors

<u>Microphone</u>

The microphone was used to generate audio recordings that were used by the experimenter and video coder when generating the objective outcome measures as well as the video-rating measures described immediately above.

The microphone was used along with the opensmile signal processing toolkit (Eyben et al., 2013) in the implementation of the computational system (as described in sections 4.2 and 5.8 of this document) to allow for the autonomous robot to identify the states of negativity and insensitivity when deciding whether or not to intervene. As described in chapters four and five and appendices A-C of this document, the paralinguistic cues used in the computational model are not defining of the problematic relationship states identified in the computational model; therefore, the video-rating measures could not be automated. The video-rating measures described above had to serve a ground truth for the computational model. Further, as described in the text, as well as in the toolkits guidelines (Eyben et al., 2013), the values output by the signal processing algorithms hold importance relative to one another with respect to how fast the individual is speaking, how much the individual is speaking in a single utterance, and how loud the individual is being. The values produced by the toolkit, however, are dependent on the environment in which they are generated. They would not reliably produce the objective measures described above.

Camera

> The camera generated visual recordings that showed full-faced view of the participant during the discussion. These visual recordings were used by the experimenter and the video coder when generating the objective measures as well as the video-rating measures described immediately above.
>
> The camera, along with the dlib (King, 2009) and OpenCV (Bradski & Kaehler, 2008) software packages, were used in the implementation of the computational system (as described in sections 4.2 and 5.8 of this document) to allow for the autonomous robot to identify the states of negativity and insensitivity when deciding whether or not to intervene. The software estimated whether the participant was oriented toward or away from the confederate in each camera frame. As described in chapters four and five and appendices A-C of this document, the orientation of the participant is not defining of the problematic relationship states identified in the computational model; therefore, the video-rating measures could not be automated. These video-rating measures had to serve as a ground truth for the accuracy of the computational model.

Empatica E4 Wristband

> Data Generated: The physiological outcome measures described above were generated with this sensor.
>
> Data Use: This data was going to be used as outcome measures for objective measures of stress in the participants and as input to the computational model to identify negative states in experiment participants. As discussed in section 5.8 of this document, however, the data from this sensor was not usable. Participants loosened the sensor during trials causing complete data lose, participants arrived to trials in different states of physiological arousal (because of running over, already being in the building, etc.). The time available to collect accurate baseline readings for each participant was not sufficient and so this data could not be used.

Experimenter Bias

As described above, the experimenter was directly involved in the generation of objective and video-rated experiment results. This is a major source of experimenter bias and precautions had to be taken to limit this bias as much as possible. Further, there were considerations as to how to detect this bias in the data generated.

Precautions to Mitigate Bias

The experimenter prepared detailed guidelines to follow when generating the objective as well as the video coding results. The experimenter took precautions to ensure strict adherence to the guidelines. When coding the objective measures, the experimenter would view videos multiple times to ensure consistency with respect to the generated measures. The video coder trained with the experimenter on six videos to reach a common understanding of the guidelines (Appendices A-C) for the video rating measures. The video coder and experimenter then coded all of the videos independently on these measures (including the six training videos). They checked their agreement every four videos to ensure that they could justify their ratings according to the guidelines and to ensure that the guidelines were applied consistently (no labels were changed during these checks). As noted in chapters five and six of this document, the interrater reliability between the experimenter and the video coder on the video rating measures was comparable to what has been seen in related literature (e.g. Jung, 2016).

The video segments were only included in the analysis if the experimenter and the video coder agreed upon the presence or absence of the problematic state. This was to ensure that the video segments that clearly demonstrate these states or clearly did not demonstrate these states are those included. The experimenter and the independent video coder had to justify their ratings using the scales in Appendices A-C. The video segments that are excluded have conflicting or ambiguous elements. Moen et al. (2001) discusses how there are moments during interactions where mediators deal with conflicting and ambiguous signals. Excluding these moments reduces opportunities for bias. In these moments, the video coders attend to disparate cues when making their decisions. It is important to ensure certain cues are not given undue weight when identifying states of weakness and alienation (Bush & Folger, 2010a; Moen et al., 2001).

Mechanisms To Detect Bias

Once the data had been generated by the experimenter and the independent video rater, there are two mechanisms that can be useful in understanding the degree to which the experimenter's bias was mitigated.

First, the data generated can be compared between participant groups to understand if there were statistical differences in the distributions from which the results were drawn. The objective and video rating measures were directly related to the experimenter's hypotheses regarding the ability of an autonomous robotic agent to support positive changes in the interactions. A lack of statistical differences between the groups is indicative that the measures arise from the same statistical distributions, which violate experimenter expectations.

The vast majority of objective measures and video rating measures, which are all described above, showed no statistically significant difference between groups. The only objective measure that showed a statistically significant difference between groups was the number of times personal feelings were explicitly verbalized when comparing participants with the intervening robot and those with the unengaged robot. The only video rating measure that showed a statistically significant difference between groups was the measure comparing the proportion of empowered statements that followed weak statements in the intervening agent and unengaged agent groups.

The second way in which experimenter bias can be detected is by considering if the measures less susceptible to this bias support or refute the findings of the coded objective measures and video coded measures. Participants completed self-report items that were analyzed directly. These items, like the video coded and coded objective measures, found that the weakness and alienation between participant groups was not significantly different. The conflict was experienced similarly for the participants in all of the experiment groups.

Participants were asked survey questions about perceptions of their partner in the interaction (the experiment confederate). The participants in the intervening robot group indicated (on self-report measures) that the experiment confederate was less willing to have a conversation of depth with them, less willing to find similarities, and less willing to acknowledge their feelings and viewpoints. The participants with the intervening agent were more cognizant of their feelings and the lack of response from the experiment confederate to these feelings. This is supportive of the video coding and objective measure results, which suggest that the participants tried to communicate their feelings more often with the intervening robot present.

REFERENCES

Adrian-Taylor, S. R., Noels, K. A., & Tischler, K. (2007). Conflict between international graduate students and faculty supervisors: Toward effective conflict prevention and management strategies. *Journal of Studies in International Education*, *11*(1), 90-117.

Antes, J. R. (2010). The Experience of Interpersonal Conflict: A Qualitative Study. *Transformative mediation: A Sourcebook. Resources for Conflict Intervention Practitioners and Programs*, Hempstead, NY: Institute for the Study of Conflict Transformation, Inc., 51-72.

Banh, A., Rea, D. J., Young, J. E., & Sharlin, E. (2015, October). Inspector Baxter: The Social Aspects of Integrating a Robot as a Quality Inspector in an Assembly Line. In *Proceedings of the 3rd International Conference on Human-Agent Interaction* (pp. 19-26). ACM.

Beisecker, A. E. (1989). The influence of a companion on the doctor-elderly patient interaction. *Health communication*, *1*(1), 55-70.

Bradley, M. M., & Lang, P. J. (2000). Measuring emotion: Behavior, feeling, and physiology. *Cognitive neuroscience of emotion*, *25*, 49-59.

Bradski, G., & Kaehler, A. (2008). *Learning OpenCV: Computer vision with the OpenCV library*. O'Reilly Media, Inc.

Brockman, J. L., Nunez, A. A., & Basu, A. (2010). Effectiveness of a conflict resolution training program in changing graduate students style of managing conflict with their faculty advisors. *Innovative Higher Education*, *35*(4), 277-293.

Burgoon, J. K. & Hale, J. L. (1984). The fundamental topoi of relational communication. *Communication Monographs*, *51*(3), 193-214.

Burgoon, J. K. & Hale, J. L. (1987). Validation and measurement of the fundamental themes of relational communication. *Communication Monographs, 54*(1), 19-41.

Bush, R. A. B. (2010). Taking self-determination seriously: The centrality of empowerment in transformative mediation. *Transformative mediation: A*

Sourcebook. Resources for Conflict Intervention Practitioners and Programs, Hempstead, NY: Institute for the Study of Conflict Transformation, Inc., 51-72.

Bush, R. A. B., & Folger, J. P. (2010a). Transformative mediation: Theoretical foundations. In Folger, J., Bush, R. A. B. and Noce, D. J. D. (Eds.) *Transformative Mediation: A Sourcebook. Resources for Conflict Intervention Practitioners and Programs*, Hempstead, NY: Institute for the Study of Conflict Transformation, Inc., 15-30.

Bush, R. A. B., & Folger, J. P. (2010b). Transformative mediation: Core Practices. In Folger, J., Bush, R. A. B. and Noce, D. J. D. (Eds.) *Transformative Mediation: A Sourcebook. Resources for Conflict Intervention Practitioners and Programs*, Hempstead, NY: Institute for the Study of Conflict Transformation, Inc., 31-50.

Bush, R. A. B., & Pope, S. G. (2002). Changing the quality of conflict interaction: The principles and practice of transformative mediation. *Pepp. Disp. Resol. LJ*, *3*, 67.

Cormier, D., Newman, G., Nakane, M., Young, J. E., & Durocher, S. (2013, August). Would you do as a robot commands? an obedience study for human-robot interaction. In *International Conference on Human-Agent Interaction*.

Costa, J., Jung, M. F., Czerwinski, M., Guimbretière, F., Le, T., & Choudhury, T. (2018, April). Regulating feelings during interpersonal conflicts by changing voice self-perception. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1-13).

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334.

Della Noce, D. J., Antes, J. R., & Saul, J. A. (2003). Identifying practice competence in transformative mediators: An interactive rating scale assessment model. *Ohio St. J. on Disp. Resol.*, *19*, 1005.

DiMicco, J. M., Pandolfo, A., & Bender, W. (2004, November). Influencing group participation with a shared display. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work* (pp. 614-623). ACM.

Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and autonomous systems*, *42*(3), 177-190.

Elangovan, A. R. (1995). Managerial third-party dispute intervention: A prescriptive model of strategy selection. *Academy of Management Review*, *20*(4), 800-830.

Eyben, F., Weninger, F., Gross, F., & Schuller, B. (2013, October). Recent developments in opensmile, the munich open-source multimedia feature extractor. In *Proceedings of the 21st ACM international conference on Multimedia* (pp. 835-838).

Feys, J. (2016). Nonparametric tests for the interaction in two-way factorial designs using R. *The R Journal*, *8*(1), 367-378.

Fischer, K. (2011, March). Interpersonal variation in understanding robots as social actors. In *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*(pp. 53-60). IEEE.

Folger, J. P. (2010). Transformative Mediation and the Courts: A Glimpse at Programs and Practice. In *Transformative Mediation: A Sourcebook. Resources for Conflict Intervention Practitioners and Programs*, J. Folger, R. A. B. Bush & D. J. D. Noce, Ed., Hempstead, NY: Institute for the Study of Conflict Transformation, Inc., 165-180.

Giner-Sorolla, R. (2012). *Judging passions: Moral emotions in persons and groups*. Psychology Press.

Haidt, J. (2003). The moral emotions. *Handbook of affective sciences*, *11*, 852-870.

Heppner, P. Paul; Wampold, Bruce; Owen, Jesse; Thompson, Mindi; Wang, Kenneth (2016). *Research Design in Counseling*. Boston, MA: Cengage Learning. p. 334. ISBN 9781305087316.

Hoffman, G., Zuckerman, O., Hirschberger, G., Luria, M., & Shani-Sherman, T. (2015, March). Design and Evaluation of a Peripheral Robotic Conversation Companion. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (pp. 3-10). ACM.

Jehn, K. A. (1995). A multimethod examination of the benefits and detriments of intragroup conflict. *Administrative science quarterly*, 256-282.

Jehn, K. A., & Mannix, E. A. (2001). The dynamic nature of conflict: A longitudinal study of intragroup conflict and group performance. *Academy of management journal*, *44*(2), 238-251.

Jorgensen, E. O., Moen, J. K., Antes, J. R., Hudson, D. T., & Hendrikson, L. H. (2001). Microfocus in mediation: The what and the how of transformative opportunities. *Designing mediation: Approaches to training and practice within a transformative framework*, 133-149.

Jung, M. F. (2016). Coupling interactions and performance: Predicting team performance from thin slices of conflict. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *23*(3), 1-32.

Jung, M. F., Martelaro, N., & Hinds, P. J. (2015, March). Using Robots to Moderate Team Conflict: The Case of Repairing Violations. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (pp. 229-236). ACM.

Kantek, F., & Gezer, N. (2009). Conflict in schools: Student nurses' conflict management styles. *Nurse education today*, *29*(1), 100-107.

King, D. E. (2009). Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, *10*(Jul), 1755-1758.

Lee, K. M., Jung, Y., Kim, J., & Kim, S. R. (2006). Are physically embodied social agents better than disembodied social agents?: The effects of physical embodiment, tactile interaction, and people's loneliness in human–robot interaction. *International journal of human-computer studies*, *64*(10), 962-973.

Levene, H. (1960). "Robust tests for equality of variances". In Ingram Olkin; Harold Hotelling; et al. (eds.). *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*. Stanford University Press. pp. 278–292.

Lewicki, R. J., Weiss, S. E., & Lewin, D. (1992). Models of conflict, negotiation and third party intervention: A review and synthesis. *Journal of organizational behavior*, *13*(3), 209-252.

Li, J. (2015). The benefit of being physically present: A survey of experimental works comparing copresent robots, telepresent robots and virtual agents. *International Journal of Human-Computer Studies*, *77*, 23-37.

Li, J., Ju, W., & Nass, C. (2015, March). Observer perception of dominance and mirroring behavior in human-robot relationships. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (pp. 133-140). ACM.

Liu, S., Helfenstein, S., & Wahlstedt, A. (2008). Social Psychology Of Persuasion Applied To Human-agent Interaction. *Human Technology: An Interdisciplinary Journal on Humans in ICT Environments*, *4*(2), 123-143.

Mallick, S. (2016). Head pose estimation using OpenCV and Dlib. https://www.learnopencv.com/head-pose-estimation-using-opencv-and-dlib. Accessed: 2018-09-02

Mann, H. B. & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, 18, 50-60.

Miller, R. G. (1966). *Simultaneous Statistical Inference*. Springer.

Moen, J. K., Hudson, D. T., Antes, J. R., Jorgensen, E. O., & Hendrikson, L. H. (2001). Identifying opportunities for empowerment and recognition in mediation. *Designing Mediation: approaches to training and practice within a transformative framework. Hempstead, NY*, 112-132.

Montemayor, R. (1983). Parents and adolescents in conflict: All families some of the time and some families most of the time. *Journal of early adolescence*, *3*(1-2), 83-103.

Moore, C. W. (2014). *The mediation process: Practical strategies for resolving conflict*. John Wiley & Sons.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of social issues*, *56*(1), 81-103.

Noce, D. J. D. (2010). Shifts, Fractures, and Supports: A Communication Perspective on Conflict Transformation. *Transformative mediation: A Sourcebook. Resources for Conflict Intervention Practitioners and Programs*, Hempstead, NY: Institute for the Study of Conflict Transformation, Inc., 145-159.

Nugent, P. S., & Broedling, L. A. (2002). Managing conflict: Third-party interventions for managers. *Academy of Management Perspectives*, *16*(1), 139-154.

Pettinati, M. J. & Arkin, R. C. (2018). A Robot to Provide Support in Stigmatizing Patient-Caregiver Relationships. *The Thirty-First International Flairs Conference*.

Pettinati, M. J., & Arkin, R. C. (2019, November). Identifying Opportunities for Relationship-Focused Robotic Interventions in Strained Hierarchical Relationships. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 297-304). IEEE.

Rahim, A., & Bonoma, T. V. (1979). Managing organizational conflict: A model for diagnosis and intervention. *Psychological reports*, *44*(3c), 1323-1344.

Rahim, A. M., Magner, N. R., & Shapiro, D. L. (2000). Do justice perceptions influence styles of handling conflict with supervisors?: What justice perceptions, precisely?. *International Journal of Conflict Management*, *11*(1), 9-31.

Retzinger, S. M. (1991). *Violent emotions: Shame and rage in marital quarrels*. Sage Publications.

Riesch, S. K., Gray, J., Hoeffs, M., Keenan, T., Ertl, T., & Mathison, K. (2003). Conflict and conflict resolution: Parent and young teen perceptions. *Journal of Pediatric Health Care*, *17*(1), 22-31.

Riesch, S. K., Jackson, N. M., & Chanchong, W. (2003). Communication approaches to parent-child conflict: young adolescence to young adult. *Journal of pediatric nursing*, *18*(4), 244-256.

Robinette, P., Li, W., Allen, R., Howard, A. M., & Wagner, A. R. (2016, March). Overtrust of robots in emergency evacuation scenarios. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 101-108). IEEE.

Roubroeks, M., Ham, J., & Midden, C. (2011). When artificial social agents try to persuade people: The role of social agency on the occurrence of psychological reactance. *International Journal of Social Robotics*, *3*(2), 155-165.

Rusbult, C. E., Martz, J. M., & Agnew, C. R. (1998). The investment model scale: Measuring commitment level, satisfaction level, quality of alternatives, and investment size. *Personal relationships*, *5*(4), 357-387.

Sabini, J., Garvey, B., & Hall, A. L. (2001). Shame and embarrassment revisited. *Personality and Social Psychology Bulletin*, *27*(1), 104-117.

Sembroski, C. E., Fraune, M. R., & Šabanović, S. (2017). He said, she said, it said: Effects of robot group membership and human authority on people's willingness to follow their instructions. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (pp. 56-61). IEEE.

Shapiro, S. S. & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3/4), 591-611.

Shen, S., Slovak, P., & Jung, M. F. (2018). Stop. I See a Conflict Happening.: A Robot Mediator for Young Children's Interpersonal Conflict Resolution. *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction. ACM.*

Simand, J. & Wright, C. C. (2005). The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Physical therapy* 85.3, 257-268.

Sirkin, D., Mok, B., Yang, S., & Ju, W. (2015, March). Mechanical ottoman: how robotic furniture offers and withdraws support. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (pp. 11-18). ACM.

Sonalkar, N., Jung, M., & Mabogunje, A. (2011). Emotion in engineering design teams. In *Emotional engineering* (pp. 311-326). Springer, London.

Steinberg, L. (2000). The family at adolescence: Transition and transformation. *Journal of Adolescent Health*, *27*(3), 170-178.

Steinberg, L. (2001). We know some things: Parent-adolescent relationships in retrospect and prospect. *Journal of research on adolescence*, *11*(1), 1-19.

Taylor, A. (2002). *The handbook of family dispute resolution mediation theory and practice*. San Francisco, CA: Jossey-Bass.

Tennent, H., Shen, S., & Jung, M. (2019, March). Micbot: A peripheral robotic object to shape conversational dynamics and team performance. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 133-142). IEEE.

Tickle-Degnen, L. (2006). Nonverbal behavior and its functions in the ecosystem of rapport. *The SAGE handbook of nonverbal communication*, 381-399.

Tickle-Degnen, L., & Lyons, K. D. (2004). Practitioners' impressions of patients with Parkinson's disease: the social ecology of the expressive mask. *Social Science & Medicine*, *58*(3), 603-614.

Tickle-Degnen, L., Zebrowitz, L. A., & Ma, H. I. (2011). Culture, gender and health care stigma: Practitioners' response to facial masking experienced by people with Parkinson's disease. *Social Science & Medicine*, *73*(1), 95-102.

Vuchinich, S., Emery, R. E., & Cassidy, J. (1988). Family members as third parties in dyadic family conflict: Strategies, alliances, and outcomes. *Child Development*, 1293-1302.

Walton, R. E. (1969). *Interpersonal peacemaking: Confrontations and third-party consultation* (Vol. 1). Reading, MA: Addison-Wesley.

Williams, A., & Nussbaum, J. (2001). *Intergenerational communication across the life span*. Mahwah, N.J.: L. Erlbaum.

Zinn, W. (1993). The empathic physician. *Archives of Internal Medicine*, *153*(3), 306-312.

Zuckerman, O., Hoffman, G., Kopelman-Rubin, D., Klomek, A. B., Shitrit, N., Amsalem, Y., & Shlomi, Y. (2016). KIP3: Robotic Companion as an External Cue to Students with ADHD. *Proceedings of the TEI '16: Tenth International Conference on Tangible Embedded and Embodied Interaction - TEI '16*, pp. 621-626.