Conversational Agents for Automated Group Meeting Facilitation

A Computational Framework for Facilitating Small Group Decision-Making Meetings

A PhD Dissertation By:

Ameneh Shamekhi

to the Faculty of the Graduate School
of the Khoury College of Computer Sciences
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
Northeastern University
Boston, Massachusetts

May 2020

Northeastern University
Khoury College of Computer Sciences

PhD Thesis Approval

Thesis Title: Conversational Agents for Automated Group Meeting Facilitation
~~A Computational Framework for Facilitating Small Group Decision-~~Making Meetings

Author: Ameneh Shamekhi

PhD Thesis Approval to complete all degree requirements for the PhD in Computer Science.

Thesis Advisor	7/21/20 *Date*
Thesis Reader	7/21/20 *Date*
Thesis Reader	7/21/20 *Date*
Thesis Reader	07/21/20 *Date*
Thesis Reader	*Date*

KHOURY COLLEGE APPROVAL:

Associate Dean for Graduate Programs	7/27/20 *Date*

COPY RECEIVED BY GRADUATE STUDENT SERVICES:

Recipient's Signature	7/28/20 *Date*

Distribution: *Once completed, this form should be scanned and attached to the front of the electronic dissertation document (page 1). An electronic version of the document can then be uploaded to the Northeastern University-UMI Website.*

ABSTRACT

Group meetings are ubiquitous, with millions of meetings held across the world every day. However, meeting quality, group performance, and outcomes are challenged by a variety of dysfunctional behaviors, unproductive social dynamics, and lack of experience in conducting efficient and productive meetings. Previous studies have shown that meeting facilitators can be advantageous in helping groups reach their goals more effectively. However, many groups do not have access to a human facilitator due to lack of resources or other barriers. In this work, I leverage advances in conversational agent technology to provide some of the functionality of human facilitators in group decision-making using an automated system.

I present a general computational framework for automated meeting facilitation that provides a real-time, co-located, synchronous, multiparty, personified interface to a group conducting a face-to-face meeting. The framework is designed to support a range of virtual and robotic embodiments of the agent, and provides facilitation functions based on group management science, including: a) management of meeting structure; b) management of participation to avoid dominance; and c) management of conflict.

Using this framework, I report on three prototypes and their evaluation to address a series of research questions. I first explored the overall acceptance of a conversational agent in the role of a meeting facilitator and compared reactions and outcomes to different agent embodiments in the meeting room, finding that embodiment improved participants' rapport with and trust of the agent as well as their perceptions of the agent's intelligence and power. Second, I developed a fully automated meeting facilitation robot that uses multimodal inputs to manage multiparty conversation, enforce meeting structure, promote time management, balance group participation,

and facilitate group decision-making processes. Results of a between-subject study showed that the robot facilitator was accepted by group members, was effective in enforcing meeting structure, and that users found it helpful in balancing group participation. Finally, I developed and evaluated conflict management strategies for the facilitation robot and evaluated them in a task setting designed to evoke task conflict. I found that participants complied with the robot's intervention, performing active listening with each other when conflicts arose. I also report on qualitative findings from interviews with study participants and a focus group, highlighting the potential role for robots in meetings of the future.

ACKNOWLEDGEMENT

Countless pieces worked together to lead me to where I am standing today. Many of those pieces, I do not even know and might never come to realize. While an appropriate acknowledgment of all of those blessings is beyond my abilities, I would like to name just a few.

First, I would like to thank my advisor Dr. Timothy Bickmore, from whom I learned invaluable lessons that not only led my research but also taught me how to design and approach research questions that are meant to address real-world problems and reduce people's pain. Your lessons will remain with me for life.

And special thanks to the members of my dissertation committee, Dr. Stacy Marcella, Dr. Vera Liao, and Dr. Lu Wang whose guidance and support were essential in every step of my research, and to produce the outcome I produced.

My internships were eye-opening experiences that helped shaping my academic character, and finding my desired filed of research. I would like to express my sincere gratitude to my great mentors at Microsoft Research and IBM Research, Dr. Mary Czerwinski, Dr. Rachel Bellamy and again Dr. Vera Liao for believing in me, their support, and their invaluable lessons all through these years. My thesis work was initiated during my second internship and I cannot thank enough the mentors who encouraged and supported my ambitious internships ideas.

In this journey I was so lucky to have the presence of my amazing friends and lab-mates. I am so thankful to my friends and colleagues at Relational Agents Group, Dina, Zessie, Shou, Stefan, Teresa and Dhaval without whom I couldn't enjoy my work as much I did during my PhD. Special thanks to my work-buddies and my nag-buddies Eve, Prasanth, and Farnaz who made these years 10 times more cheerful for me.

I would also like to thank Ha Trinh who was very inspiring during important years of my PhD and I learnt a great deal from, and Lin Shi who kindly provided everything I needed, and Juan Fernandez for all his technical support.

And finally my gratitude goes to my husband, Mohammad, who patiently supported me and motivated me to proceed in this path, my father in Tehran whose support and calm words helped me handle my stressful days, and my brother, sister, sister in law and nephews who are the bright spots of happiness and joy thinking of whom has brighten up my path.

Table of content

Chapter 1: INTRODUCTION

At corporations and organizations, meetings are the primary means of communication, and on average, a third of employees' time is spent in meetings [2], [3]. A 1996 research study estimated that over 25 million meetings were held daily in the United States, and over 85 million worldwide [2]. Extending the curve to more recent times results in between 39 to 56 million meetings daily in the US alone [2]. Teams often collaborate face-to-face in order to make decisions that reflect their common interests and goals in the form of group meetings. A meeting is formally defined as a focused interaction of a group of people with a common purpose [164]. Group meetings enable employees to view a problem from different perspectives, get exposure to a diverse set of insights and ideas that are not accessible otherwise, and reach consensus on a solution. However, there are many challenges in group meetings, endangering the productivity of these meetings.

Despite the many advantages of group meetings, these meetings create a complex network of social interactions that can consume an organization's time and energy if not managed. Even in a small group, collaboration amongst individuals with different interests, backgrounds, personalities, and cultures can create a complicated environment, potentially impacting group performance and outcome. As a result, numerous researchers in management sciences and social psychology have explored the dynamics, and different interactions, roles, and tasks people take in meetings to classify phenomena that can affect group performance.

Group dynamics such as social influence, social loafing [4], and groupthink [5] (see Chapter 2.1) have been shown to complicate the decision-making process and impact a group's performance. Inadequate management of meetings, such as the lack of an agenda (or having a vague or redundant agenda), deviation from an existing agenda, hidden agenda, imbalanced participation, and distraction, are other factors that can result in unproductive meetings[80]. Lacking an enforced structure to the meeting has caused 'getting off the subject', 'no goal or agenda', and 'too lengthy' to be the top three reported problems of meetings [164]. This highlights the positive impact of employing structure management mechanisms on group satisfaction and productivity [103].

In addition to poor meeting structure, intragroup conflicts have been shown to be another factor that can affect teamwork negatively [149]. Conflict and disagreement are among the most common challenges in a collaborative work environment, which can impact the group performance both positively and negatively. Conflict in small groups occurs when the interests and beliefs of group members do not match, which leads to the group not being able to reach a consensus on a decision. Intragroup conflicts refer to disagreements between the members of the group. Jehn and Mannix identified three types of intragroup conflicts: relation, task, and process conflict, which can all be impacted by several factors such as power dynamics in the group, the members' relations, emotions and the minority-majority structure in small groups [95]. Although strong conflicts in groups can hardly be helpful, a moderate level of disagreement may be constructive by providing attendees with the opportunity to see aspects of an object from different points of view [95]. Regardless of the level of disagreement, the most agreed upon opinion about intragroup conflicts is that they can be very harmful if left unmanaged and unresolved [12, 76]. Therefore active conflict management is essential for any group to maintain group satisfaction and performance, while ignoring disagreements can impair member relationships and group performance [51].

As a result of these challenges, it is hard to keep a group meeting productive. Many reviews and surveys revealed that even though meetings take a considerable amount of employees' time, they are evaluated as unproductive and dissatisfying [164]. A report by Harvard Business Review in 2017 revealed that more than 70% of senior managers "considered meetings unproductive and inefficient" [155]. A survey in 2014 also estimated the cost of unproductive meetings at more than $37 billion per year [2][164]. These estimates have created some debates about whether meeting as a *group* necessarily results in a more effective problem-solving process and better outcomes compared to *individual* approaches [6][7]. In the past 60 years, many researchers in social psychology [8], business sciences [9][10], and computer-supported cooperative work [11] have explored ways to improve the productivity and performance of collaborative decision-making processes.

1.1 Conversational Agents as Automated Meeting Facilitators

Many studies have suggested that providing structured communication and decision-making procedures are essential interventions to promote constructive behaviors and mitigate dysfunctional group dynamics [6]. Building upon the findings of prior research, in this work, I seek to leverage conversational agent technology to support group meetings by enforcing structure to the meeting and facilitating decision-making processes.

Professional human facilitators have been shown to be effective in performing these functions in group decision-making meetings. For example, Westley and Waters have shown that group facilitators are effective at addressing many of the common problems in meetings, such as not following the meeting structure (or lacking a structure), intragroup conflict, and domination of the discussion by one participant [20]. Viller also showed that a group facilitator could greatly improve meeting quality by enforcing a structure, eliciting equal participation, and managing conflicts [11].

Despite the potential benefits of group facilitators, the high cost of hiring professional consultants or training in-house facilitators alongside with logistical barriers prevents many organizations from using them. Moreover human facilitators may not have access to the knowledge and information related to the team's task and familiarizing them with the team's technical domain is time-consuming and costly. An automated group facilitation system may be able to perform many of the functions of a human group facilitator at a fraction of the cost. In this work, I use the previous literature on group meeting dynamics, group performance, and group facilitation to design a computer-supported group facilitation system.

In my group facilitation system, a conversational agent plays the role of a Virtual Facilitator to provide the most natural and intuitive interaction with group members and enhance the performance and quality of a group meeting. Conversational Agents (CAs)—virtual or robotic—have several capabilities that can make them effective meeting facilitators, including: the ability to interact with human collaborators in natural language and nonverbal conversational modalities, the ability to hold a non-judgmental and neutral point-of-view, and access to information resources. . The CA can use speech to convey the instructions in an optimized manner and interact with group members, and group members can speak back to the agent with no need for training. Several researchers have investigated the design and development of CAs and shown the effectiveness of them in one-on-one interactions in various domains, such as healthcare [22] and education [23]. However, there is a lack of understanding of how CA's can function in a group setting, and this thesis work aims to fill this gap by exploring ways in which CAs can facilitate a group decision-making session.

In this document, I use the term 'Conversational Agent' as a general term referring to any conversational interface with human-like personification. A conversational agent could have various forms of embodiment. It could have no embodiment (e.g., a voice-only character), or be embodied in the form of an animated character on a screen, or have a physical embodiment like a robot. Regarding the conversational agents for the group facilitation, I will use the terms "virtual facilitator" and "group facilitation robot/agent" interchangeably to refer to a conversational robot/agent that facilitates a group meeting.

1.2 Problem Domain: Decision-Making Meetings

In order to develop a clear framework and methodology, in my work I focus on small group decision-making meetings that are a ubiquitous example of group meetings. Group decision making meetings are a common meeting scenario in which employees can view a problem from different perspectives, get exposure to a diverse set of insights and ideas that are not accessible otherwise, and reach consensus on a solution.

A very common example of group decision-making meetings is a *hiring decision meeting*. In a hiring meeting, two or more co-located individuals review a set of resumes to decide on the one best candidate to hire. Previous research also used this setting to study group decision processes. In my research I use 'hiring meeting' as a prevalent type of group decision-making sessions, to clearly define procedures and facilitation provided by my conversational agent-based group facilitation system. However, the methods and technologies developed are immediately applicable to a much broader range of group decision-making problems.

In all types of group decision-making scenarios, including hiring sessions, the 'decision-making processes' in meetings are one metric used to assess meeting performance. Two factors have been frequently studied regarding the decision-making process in prior research: *structure management* and *conflict resolution* [12], [13] that could go a long way towards improving meeting quality if done properly. *Structure management* involves following a meeting agenda in a structured way, and several studies have shown that following a structured decision-making process leads to higher satisfaction, better decisions, and more creative solutions [14][12], [15]. *Conflict resolution* involves detecting and mitigating intragroup conflicts. Studies have demonstrated that intragroup conflict is a major source of meeting inefficiency [16], and unmanaged conflict can deeply impair team performance. Thus, structure and conflict management are two important factors that a team should consider to further optimize their performance.

1.3 Research Questions

In this work, I propose an automated approach to address some of the known challenges in a group decision-making meeting, including lack of structure, individuals' dominance, and intragroup conflict, using a conversational agent as a group facilitator. I develop a Virtual Facilitator that sits in the room with the meeting participants and enforces a meeting structure, monitors and balances members' participation, and detects and manages conflicts. A virtual facilitator can see, listen to, and partially understand all group members and provides just-in-time instructions to assist them.

I seek to address the following research questions in my dissertation work:

RQ0- Will members of a face-to-face decision-making meeting accept a CA in the role of a group facilitator?

RQ1- What would information workers expect from a CA at their workplace? 5.a) what role and features would they prefer for the robot to hold?

RQ2- What is the appropriate embodiment for a virtual facilitator in a group setting?

RQ3- To what extent do the members of a group follow the virtual facilitator's instructions and recommendations in the context of group decision-making?

RQ4- How can a virtual facilitator impose and enforce a **structure** on a group decision-making activity, and ensure that all participants have an opportunity to be heard?

RQ5- Can a virtual facilitator improve conflict management in a group decision-making setting?

Through an incremental development process, I report the development of three prototypes for group facilitation CAs, and a series of user studies and focus groups to evaluate the effectiveness of such agents in improving people's group meeting experiences and managing multiparty interactions, and to answer the research questions above.

I First develop a preliminary prototype to study the feasibility, explore the use of CA as group facilitators in the group setting. Using this semi-automated prototype, I also seek to find out the right embodiment for a virtual facilitator via a Wizard of Oz experiment. As the second prototype I develop a fully automated group facilitation system using multimodal sensor inputs, that facilitates a groups decision making session by enforcing meeting structure, managing participation, and facilitating decision-making processes. I conduct a user-study to evaluate this automated system. Then I focus on intragroup conflicts as a common challenge in group meetings, and design a data-driven model to detect disagreements in a group conversation. I also study different conflict management strategies delivered by a virtual group facilitator in a Wizard of Oz study. Finally, I conduct a focus group with real information workers to find out what employees would need and expect from a virtual facilitator at real work-setting.

My goal in this research is to understand the right design for a CA in the role of a group facilitator and to identify the expectations from such a CA. In large part, my goal is to inform the design of social conversational systems for group facilitation in workplaces. To this end, I have also investigated what existing mental models participants hold for a virtual facilitator, and describe how working with a facilitation agent may be similar to or different from working with a human moderator in a group meeting.

This work contributes to the fields of HCI and CSCW by providing a comprehensive account of the impact of CAs in a group collaboration setting. My dissertation project also contributes to the field of group decision support systems by definign different types of facilitation (e.g. social, meeting, and decision-making facilitation) and outlining the required functions for each facilitation type to be provided by an automated system. One of the primary contributions of this work is the design of a computational framework that can drive a CA in a multiparty interaction to provide the main functions outlined above: facilitating meeting structure, managing group participant, and detecting intragroup conflicts in group meetings, with the goal of improving the overall quality of the group outcome. Last but not least, I studied different strategies that could be taken by a conversational system in a group decision-making setting to manage disagreement. Using the findings of my studies I created a design guideline for conversational interfaces for group meeting facilitation (section 9.2.5) and highlight the technical and design requirements of such systems in real work settings. The framework and guideline can be used by researchers to further investigate the design and applications of automated group facilitation systems.

1.4 Dissertation Outline

This dissertation contains eight chapters that are organized as follows.

In Chapter 2, I present previous related work in multiple areas, including work on the use of robots in groups and research in group behavior literature. I also review theoretical concepts that will be used in later chapters.

In Chapter 3, I describe a framework for a Conversational System for Group Meeting Facilitation. I also present the rationale behind the design of specific functions and capabilities of the group facilitation system. Subsequently, I declare the technical requirements of such a system.

In Chapter 4, I describe the first prototype of the group facilitation system that I developed to assess the overall acceptance and feasibility of my group facilitation system. Through a wizard of Oz study, I investigated the right embodiment for the group facilitation agent.

In Chapter 5, I build upon the findings of the study reported in chapter 4 to justify my choice of having a physical robot as a group facilitator. I then present the technical effort to develop a fully automated group facilitation robot, followed by reporting results of an evaluation study.

In Chapter 6, I present a machine learning model to detect disagreement in group conversations. I describe the datasets used to train the models, as well as the data processing and NLP techniques I utilized to train a disagreement detection model.

In Chapter 7, I propose a disagreement management strategy to be delivered by the group facilitation robot. In a wizard of Oz study, I evaluate the effects of active disagreement management interventions by comparing them with passive interventions to resolve a disagreement in a group discussion.

In Chapter 8, I report the qualitative findings from the interviews I conducted with participants in the user study and the focus group I conducted with real work-place employees.

In Chapter 9, I present an analysis and discussion of the findings of my studies, highlight design guidelines and implications for group facilitation systems, and discuss directions for future research.

Chapter 2: RELATED WORK

2.1 Small Group Dynamics and Challenges

Different aspects of small groups; from the group structure and ecology to the group performance have been investigated by social psychology researchers over the past 70 years [172]. Social psychologists have studied the member roles and tasks to classify the individual behaviors in group structure. Back in 1948, Benne and Sheats proposed a schema for functional roles in small groups introducing 26 roles for members of a group, under three categories: task roles (e.g., initiator, information seeker), social roles (e.g., harmonizer, follower) and individual roles (e.g., aggressor, dominator) [66]. Drawing from Benne's work, Bale has proposed a behavioral coding systems for analyzing the interactions among individuals in a small group [9]. Bales' Interaction Process Analysis (IPA) model contains 12 behaviors related to social and task roles, under 4 categories: integrative/positive reaction, attempt to answers, questions, and negative reactions.

Regarding the group performance, researchers suggest that the group's *goal* and *task* are among the main defining factors of any group, in order to perform. The group performance is usually examined in a problem-solving or a decision-making setting where multiple individuals collaborate to make a decision out of several alternative options. In small groups, the performance, the decision-making process and the outcome could be potentially influenced by several factors including the social influence among members, groupthink, group polarization, and shared information biased. In the following paragraph I will describe how the group dynamics and challenges such as meeting structure and conflict that could impact the group performance.

"Social impact theory" by Latane describes the social impact as a function of three social forces: strength (power) of the source of impact (e.g., the age and social class), the immediacy of the event, and the number of individuals in the target group [116]. This theory characterizes the complex interactions among different social factor in groups which together determine how individuals in groups are impacted by the other members. It also describes how lacking an agenda, besides having derailers and dominants in a meeting, often result in frustration and going off-track in meetings.

Conflict and disagreement are among the most common challenges in a collaborative work environment which could impact the group performance both positively and negatively. Researchers have identified three types of Intragroup conflicts: relation, task, and process conflict [95]. Jehn studied the structure of 105 work groups and concluded that conflicts may benefit the group depending on "the type of conflict and the structure of the group in terms of task type, task interdependence, and group norms." I will review the previous work on conflict and conflict management frameworks later in this section.

Last but not least, Groupthink is another major challenge when a group of individuals collaborate to make a decision. Groupthink was introduced by Irving Janis for the first time in 1971 when he was trying to explain the reason behind some of the wrong foreign policy decisions made by a group of noble people [82]. This phenomenon refers to malfunctioning of decision-makers as they collaborate in a group because of a high tendency to keep the ingroup harmony and cohesion which may result in them being less critical of their leaders and colleagues' opinions to minimize the conflict in the group.

In a review paper about Group Support Systems, Nunamaker et al. examined different factors that could affect a teamwork negatively including domination, lack of focus, hidden agenda, conflict, distractions and poor planning [149]. They reviewed the Team Theory and proposed a model for group-ware evaluation. Based on Team Theory the cognitive load of individuals in group activities is divided among three processes: communication, deliberation, and information access [149]. Nunemaker's groupwear grid suggests that any technology intervention should aim to support one or more of these three cognitive processes to improve the productivity of the group.

2.2 Conflict and Conflict Management Strategies in groups

In this section, I first describe the intragroup conflict, different types of it, and the ways in which it affects the teamwork and members' emotions. I then review the conflict management theories and frameworks.

"Intragroup conflict is defined as the degree to which members have real or perceived incompatible goals or interests." [198] Research on group conflict in the past two decades has introduced four types of conflicts resulting from different forms of disagreement among group members: task conflict, relationship conflict, process conflict, and status conflict. The first two types of conflict were introduced in a series of classical studies around 1995 based on work-related and personal conflicts [9] [24]. Task conflict, as its name suggests, results from disagreement over ideas, and perspective about the group task, while relationship conflicts refers to interpersonal clashes and tensions [198]. The third form of conflict (process conflict) is about "how" responsibilities are assigned to members for task completions[94]. Status conflict, defined by Bendersk is about the hierarchical power relationships in organizations where people may disagree to one another to advocate their social positions in the power hierarchy [76].

Task conflict is associated with both negative and positive effects on group members in the past research. This type of conflict can impair the group satisfaction and emotion, when members feel they do not receive a positive assessment from their teammates [42, 201]. Group performance could be also affected by task conflict, as disagreement in group can create extra distraction and cognitive load (e.g. [31]). However, effects of task conflict on the group are not always negative. De Wit has conducted a meta-analysis on 116 empirical studies to investigate the effects of four types of conflict on group performance. Results of his investigation on the total of 8880 groups interestingly indicated that task conflict is not always harmful, but it can even improve the group performance in the absence of *relationship conflict* [198].This positive correlation between task conflict and performance can be described by the fact that different and even conflicting ideas in groups provides the members with the chance to view each topic from different perspectives, and consequently increase their understanding of the task [4]. However the three other forms of conflict all were shown to be negatively correlated with the group performance [4] Task conflict is the most common type of conflict in groups, and in this research I focus on that in the newly formed small groups with no previous relations, to simplify the scope of the project. Related to the domain of my research, in the next paragraph I delve into depth in task conflicts to review the influencing factors on that.

As mentioned earlier, several studies showed that task conflict can only help the group when there is no personal (relational) conflict among the members [199]. For example, people with higher level of emotional stability, openness [25] as well as the emotion regulation [48] in a group are more likely to benefit from task conflict, by decoupling task and relations. Team atmosphere is also important; for example, when the level of trust is high among members of a team, they are less likely to connect task conflict with emotions and personal relations.

Previous studies have also examined effects of different conflict management behaviors, and task types on how task conflict affects the group outcome and cohesion. The results however, are somewhat mixed, as Tekleab and Quigly confirmed the moderating effect of conflict management on task and relationship conflict, but they showed that *less strict* conflict management approaches inflame the benefits of *task* conflicts for group by increasing the group cohesion, (for relationship conflict, more strict conflict management styles were shown to be more effective on cohesion, though) [178], while [51] suggested that active management of conflicts and engaging members in making consensus will increase the positive effects of task conflict. In a similar vein, studies on the task types suggested mixed results. Several studies showed task conflict is more positive for nonroutine tasks such as creative brainstorming, versus routine tasks (e.g. decision making) [93] [35]. But DeWit did not find any difference in the relationship between task conflict and group performance across mixed project, creativity and decision making, and planning tasks [198].

Overall, task conflict is the only type of conflict that can potentially benefit the group outcome [198], and this is why many companies cherish this type of conflict in their culture, and researchers seek to investigate the right balance and conditions to maximize the benefits of task conflicts in groups. Concluded from past research on task conflicts, it is clear that "the effective management of conflict is critical" to optimize its effect [76] , otherwise the group cohesion and performance could be harmed by any unmanaged conflict. The two main team development theories also [73] [185] confirms that conflict management in the middle stage (storming stage-Tuckman's model) is crucial for team performance. In the rest of this section I review the literature on conflict management styles and framework.

Since the recognition of conflict as one of the impacting factors in group decision making, researcher have investigated the conflict management models with the goal of maximizing the positive outcomes of conflict and decreasing the negative outcomes. The history of research on conflict management styles and techniques can be traced back to the 1960s when Blake and Mouton proposed five classes for interpersonal conflict management focusing on "concerns for people" and "concerns for task": forcing, withdrawing, smoothing, compromising, and problem solving. A decade later, researchers began to focus more on the concerns of the *two sides* of a conflict. This approach resulted in several conflict management models in 1970s including Thomas-Kilmann Conflict Mode which introduces five modes along two dimensions: "(1) assertiveness, the extent to which the person attempts to satisfy his own concerns, and (2) cooperativeness, the extent to which the person attempts to satisfy the other person's concerns." The five conflict handling behaviors in their model are: competing, accommodating, avoiding, collaborating and compromising. Similar models have been developed with the same approach [114], [51] [160]. Rahim proposed a meta-model with five approaches for conflict handling in groups: *Integration* which involves openness, encouraging to exchange information, and offering wide range of alternatives, and examining to solve the problem. *Obliging* refers to attempts to minimize the differences and highlight the commonalities to satisfy the concern of the other party. *Dominating* approach in which one side of the conflict only focuses to win his or her objective and, consequently often ignores the expectations of the other party. The fourth approach is *Avoiding* when a party fails to satisfy his or her own concern as well as the concern of the other party. Compromising is the last approach and involves give-and-take whereby both parties give up something to make a mutually acceptable decision [160].

Conflict management can be also examined from a facilitator or leader point of view. Greer suggests three tips for managers to handle conflict in their groups: First, understand the topic of conflict, second, identify people motivation of engaging in the conflict. Finally, monitor the emotional status of the group and every individual to take a proper action when anyone becomes emotional. Poole has also mentioned actions for effective conflict management that can be incorporated in GDSS. The conditions include: stay focused on the problem rather than emotional and personal relations, considering a wide range of alternative solutions, creating a cooperative atmosphere, and enforcing an organized and orderly process. He also declared that enforcing the final decision to the group or artificial conflict-reducing devices can be very harmful for the conflict resolution in groups [158]

2.3 Group Facilitation

Much research has suggested that having a group facilitator can improve the group performance and decision making effectiveness, by enforcing a structured decision making process [21] [188] [134]. Thus, a number of Group Decision Support Systems (GDSSs) have been designed to support or even replace some of a human facilitator's roles. For example, to tackle lack of leadership in distributed groups, Farnham et al. customized a chat system to enforce the structure of the communication and the group decision-making process. This system is designed specifically for a hiring decision making scenario, where the user groups need to follow the three steps to make a decision of a best job candidate: 1) discuss the problem; 2) fully explore each alternative; and 3) rank the alternatives. The evaluation user study proved that the participant groups using the system were more likely to reach consensus, made higher quality decisions and showed a greater recall of discussions [61].

In 1991, Stephen Viller examined the role and responsibilities of a group facilitator in computer supported cooperative work (CSCW), and discussed the extent to which this role should be automated in the group setting. He described how the importance of facilitator's role changes over the group’s lifecycle, from more central role in the beginning, to providing macro interventions from the background in the middle stage, and again becoming a key role integrating the ideas and wrapping up the meeting. In that regard, Brochet also suggested 20 responsibilities for meeting moderators based on the temporal stage of the meeting: a) Successful beginnings; b) Nurturing the introductory stages; c) Maintaining the mature conference; and d) Wrapping up the conference [28]. Regarding the membership status, the facilitator can be either separate, included but different in the group, or totally a member of the group. Viller also reviews five common challenging scenarios in meeting and possible interventions for the facilitator (interpretation or direct intervention). As identified by Viller the five generic meeting problem that can be improved by a group facilitator, include lack of structure, domination, conflict, lack of progress record, lack of engagement [188]. Nunamaker et al. have also declared four functions [22] for an electronic facilitator: “(1) provides technical support by initiating and terminating specific software tools; (2) chairs the meeting, maintaining and updating the agenda; (3) assists in agenda planning; and finally (4) provides organizational continuity, setting rules and maintaining an organizational repository.”

2.4 Computer-supported Systems for Group Facilitation

There has been numerous research efforts attempting to leverage Artificial Intelligence and pervasive computing to improve group performance. Researchers have developed systems to facilitate group communication, support the decision-making process, and even manage the ingroup conflict in the workplace. In this section I review previous work on computer-mediated systems to support collaborative work in groups.

2.4.1 Group Decision Support System

Teams often sit together to make decisions that reflect common interest and goals, and researchers have built systems to support that. The term group decision support systems (GDSSs) refers to a broad collection of computer technologies that support problem formulation and solution in group meetings, including electronic messages, visualization, data analytic tools and group recommender systems. In a 1987 foundational paper, DeSanctics and Gallupe [54] defined a taxonomy of three levels of GDSSs. Level 1 GDSSs provide technical features aimed at removing common communication barriers, such as large displays for sharing ideas, preference solicitation, and anonymous input. Level 2 GDSSs provide decision modeling techniques aimed at reducing uncertainty in group decision processes, often providing automated planning and analytical tools. Level 3 GDSSs are characterized by machine-induced communication patterns with formalized procedural rules (e.g., parliamentary procedure). Within the HCI community, the studies of GDSSs can be traced back to the 1980s. Poole *et al.* [157] conducted a study to systematically understand the impact of GDSSs on the communication process of groups engaging in conflict management. They compared baseline groups, to both manual and computer supported groups. The computer supported group used a Level 1 GDSS, with a display of problem definition, criteria for evaluation, alternative solutions and the current voting status of the group, while the manual supported group had only a pen and paper version of the GDSS. Their work suggested seven types of benefits that GDSSs could provide for group conflict management, including improving expression of affect, emphasizing expressed positions, de-emphasizing personal relations, equalizing member participation, making process clearer, influencing course of the conflict, and stimulating exploration of alternatives.

2.4.1.1 Intelligent Group Decision Support System:

Intelligent Decision Support Systems (IDSS) refer to systems that adopt AI techniques and computing technologies to provide decision makers with an additional support. The services provided by an IDSS include; data mining and information retrieval to acquire a knowledge from existing data, representation and reasoning based on the acquired knowledge, identifying problems and suggesting possible solutions [187]. Expert systems are a well-known type of IDSS that simulate the decision-making process by a human expert [92]. Related to the topic of my work, in the following paragraphs, I review previous research on IDSSs for supporting collaborative work in groups.

Collaborative group work requires members to be able to communicate, share document and information, brainstorm and vote [187]. As identified in a framework by DeSanctis and Gallupe [54] Group Support Systems (GSS) in a decision room, web-based GSS, multimedia presentation and document sharing are some of the support technologies for group meetings happening in the same time and same place [54].

2.4.2 Smart Group Meeting Rooms

Smart Meeting Rooms (a.k.a Intelligent Meeting Rooms) are usually equipped with different sensors, multiple cameras and microphones to capture details of a meeting and detect contextual information [131]. Yu and Nakamura have explored the existing research and technologies for Smart Meeting Rooms (SMR), and presented an overview of different attributes of a SMR. They introduced a three-level architecture for SMRs based on the existing work. The first level is “meeting capture” containing the input devices such as camera, microphone and other sensors to record video, audio and other contextual data. The second layer is “meeting recognition” that is responsible for providing low-level information from the recorded data. The input data are analyzed in this layer, and output information such as; the speaking participant, transcripts and summary of the meeting, and recognized activities (e.g., gaze toward the board, or presenting). Another potential output of this level is hot spot recognition. Gatica-Perez et al. used audio-visual cues to detect the level of group interest and engagement [68]. In another example, Wrede and Shriberg adopted prosodic cues to detect hot spots in meetings and found that F0 and energy are able to predict ‘involved’ and ‘non-involved’ utterances in meetings. The last level in this architecture, is “semantic processing” layer which gets the output of recognition layer, and prepare understandable information for the final user. Annotating, indexing and browsing are the techniques used in this layer, when the first two classify the raw data by creating labels for meeting segments, and the later technique enables the end user to interact with the meeting data by providing a visualization of the extracted features of the meeting [202]. Figure 1 shows the generic architecture with basic modules of a smart meeting system as depicted in [60].

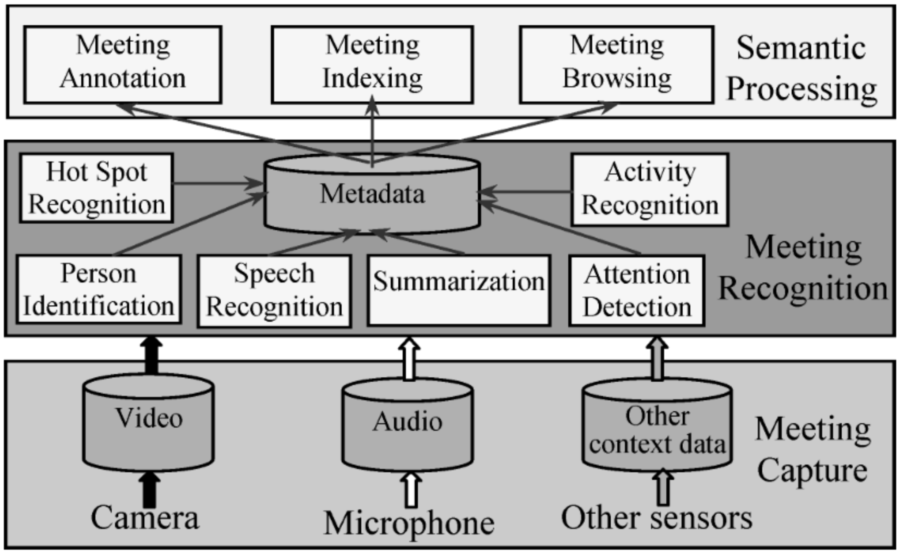


Figure 1. Generic architecture of a smart meeting system [60]

Development of such SMRs can be traced back to late 1990s when Alex Waibel at Carnegie Mellon and his colleagues introduced SMaRT: The Smart Meeting Room Task with the goal of keeping human focused on their task instead of the tool. SMaRT can identify participants and their emotions and activities such as entering the room, and speaking, events such as phone ringing and doors opening, and their focus of attention. They also developed a rich meeting browser module which enables the user to easily review different aspects of the meeting such as summaries, records and conclusions [191].

EasyMeeting is a smart meeting room developed by Chen et al. that provide relevant information to facilitate participants' tasks based their contextual needs. Design of this SMR is focused on providing context-aware support to the participants while protecting their privacy. EasyMeeting was built on a pervasive computing system and a context broker architecture that acquires and maintains a context for each entity in the meeting to help the system decide on which service it should provide. The services include predefined speech recognition, verbal presentation command, and music and lighting control [37]. In the same line, there have been several studies focused on the group facilitation services which I review in the next subsection.

Besides GDSSs and Smart Meeting Rooms, researchers have also developed systems specifically to support and facilitate group activities. Research in this area is both focused on the detection of group status and individual group-related behavior, and providing services such as floor management and meeting summarization.

2.4.3 Automated recognition of decisions and conflicts in group decision making sessions

Several studies have leveraged AI techniques to detect specific events in a meeting such as decision-making moments or disagreement between members. For example Kim and Rudin applied supervised and unsupervised algorithms on AMI dataset[1] to predict when important actions related to a decision take place in a meeting and found SVM-based models and Naive-Bayes Gaussian were able to predict key decision times with F-measure greater than 0.84 [107]. Many researchers attempted to detect conflict behavior in group conversations. Germesin and Wilson explored different sets of features that can better detect agreement and disagreement utterances in a meeting. They included lexica, prosodic and structural features in two machine learning models (a decision tree and a conditional random field (CRF)) and found CRF model without in higher level information labeling resulted in more precise prediction of (dis)agreement [70]. Several papers also investigated the automatic detection of (dis)agreement in discourse based on nonverbal behaviors. Bousmalis et al. describe the nonverbal behavior cues of agreement and disagreement group conversation. They also reviewed several available databases and 8 papers that used machine learning classifiers to detect (dis)agreement in discourse [23].

Computer-aided floor management programs have also received great attention from researchers. Chen et al, presented a new schema to predict floor control related behaviors, by annotating verbal (speech) and nonverbal (gestures and gaze) behaviors in two meeting sessions from VACE multi-modal meeting corpus. They introduced four transitional behaviors related to floor control: change, overlap, stop and self-select. They also found that the meeting leader or facilitator plays an important role in floor transitions [38]. In another study Chen et al, applied machine learning algorithms to automatically detect floor control shifts based on verbal and nonverbal features [39].

---

[1] AMI is multimodal corpus of group interactions contains more than 100 hours of meeting recordings. The recordings are annotated with several structural and behavioral tags such as dialog acts, adjacency pairs, arguments (positive, negative, uncertain) and hand gestures. [See section 6.3 for more details]

2.4.5 Robots and Conversational Agents for Human Groups

In 1996 Nass and Moon initiated the research on the feasibility of having computers as teammates and found that people affiliate with a computer as team when they are told that their evaluation is interdependence with the computer [77]. Advancements in robotics technology caused an increasing trend in the employment of robots in groups to support the teamwork in a wide range of tasks from invasive surgeries to space missions in recent years. Several researchers have also explored how the presence of robots in teams can affect the team performance, while robots' impact on the team interaction has received less attention. Malte Jung has hold two workshops at CHI and CSCW conferences and discussed the robot's social impact in groups and how placements of robots can influence the group dynamics in organizations. Earlier in 2015, Malte and his colleagues also explored using robots to moderate team conflicts through a Wizard-of-Oz study in which a robot attempted to repair interpersonal violations while a group of 3 people were solving a problem [100]. In [98] Jung reviewed previous work on the role of emotion and affect in Human-Robot-Interaction and introduced affective grounding as one of the essential pillar in coordinating HRI. He proposed that building shared understanding of team member's emotion and emotion regulation in a group is important for developing a relationship in HRI.

In regard to group failitation, Yoichi Matsuyama has conducted a series of work investigating how robots can *facilitate* the interaction among individuals in a small group, specially through engaging all participants in the conversation and managing the floor. They proposed a facilitation framework to balance the participants' engagement in the conversation, based on the functional roles in small groups introduced by Benne and Sheats [66], and the participation structure by Clark [126].

Groom and Nass have revisited the human-computer collaboration by exploring "robots" as teammates in 2007. They first cited the features of successful team such as sharing a common goal, trust in each other, share mental models and viewing interdependence as positive. Then they discussed how robots fails to meet many of these requirements and therefore are not qualified to be teammates. They stated that robots are unable to establish trust, have mental model and subjugate individual needs for the benefit of team, thus they cannot be "teammates", and most of the human-robots team introduced in previous research are in fact "human-robot organizational structures". After introducing a benchmark for human-robot teams, they recommended researchers to focus robots' advantages and special abilities in groups, instead of trying to "make robots into people" [77]. In another word researchers should study how robots can assist and complement human tasks in teams, rather than trying to replace human roles in teams. They concluded by proposing a 4-question benchmark for researchers to determine the best model for HRI.

Conversational agents (CAs) are computer systems designed to interact with users through natural conversations. Numerous researchers have discussed the design, development, and evaluation of CAs that are designed for for 1) Task-oriented personal assistance services via short-term interactions; or 2) Goal-oriented interactions to provide training or counseling content to help users achieve goals, or promote behaviors over long-term intervention. I consider a group facilitation agent to belong to the latter category. Studies have explored the potential use of goal-oriented CAs in various domains, such as healthcare and education. Studies showed these agents can improve the task outcomes in various domains such as healthcare and education. For example, Graesser et al. developed AutoTutor, an animated pedagogical agent that interacts with a learner through natural conversation [74] and exhibited the agent significantly improved students' learning by engaging them in an interactive conversation. Although the research on CAs mostly was focused on individual interactions, there is also a body of literature on virtual and robotic agents in multiparty interaction [19, 124]. For example, Kumar and Rose proposed a new architecture for building pedagogical agents to enable them to handle complex interaction dynamics in a multi-learner collaborative environment [19]. Bohus and Horvitz also proposed a model to represent a CA's turn-taking behaviors by gaze, gestures and speech. Such a model could smooth the CA's speech flow in a multi-party conversation. They used a voice activity detector and a basic turn-taking policy. It takes the floor when someone asks her a question, or when the floor is released to someone else, but that person does not take the floor [19].

Summary of Related Work and My Work Contributions

Reviewing different challenges influencing group performance and previous research on how human facilitators and group support systems may be able to address some of these challenges, point to the great potential of employing intelligent systems for group facilitation. While there has been a significant amount of research in related areas, most of them were focused on information (e.g., Information search and retrieval) and documentation facilitation (e.g., information sharing and transformation, recording the meetings, integrating equipment) [105], rather than relational facilitation and group interactions which are the targets of this work. My research is unique in that it explores the use of a humanoid CA in the role of a group meeting facilitator. The specific functions—including meeting structure management, participation management, and conflict resolution—also represent novel contributions. Whether people accept and follow a computer agent as a facilitator in the group, and whether the agent can accurately detect and effectively manage the conflicts, invite empirical research that I sought to address in my research.

This work aims to explore CAs' role in the less-studied context of group decision making. In this work, I describe a conversational agent that supports group decision making tasks, with capabilities to enforce a more structured meeting process. In contrast to [61], I focus on group decision making that happens at the same time and in the same place. According to McGrath's typology [184], such a task covers four modes— inception, problem solving, conflict resolution, and execution, and a successful collaboration should fulfill the requirements of three functions—production, well-being, and member support. Unlike the passive GDSS in [61], my agent system is designed to imitate a human facilitator by providing pro-active and humanized facilitation. Previouse research has shown that GDSSs are most effective when used with a human facilitator in a meeting [6], and my group facilitation system aims to offer best of both worlds by providing some of GDSS's functionalities via an interactive human-like computer agent. My system leverages the human-likeness of a conversational agents besides the unique capabilities of virtual agents (such as holding a natural point of view and access to information) to facilitate a group decision-making task. My agent should be considered a Level 1 GDSS, according to [54] (2.4.1). However, my interactive agent systems are able to support not only the decision-making process (production), but also positively influence the interactions of group members (well-being and member support).

Chapter 3: GROUP FACILITATION SYSTEM FRAMEWORK

The previous work on small group dynamics and group facilitation –described in the last section– demonstrated the advantages of having a facilitator in a group decision-making setting. However, the high costs of training and hiring professional facilitators prevents many groups from using them. The recent advances in conversational systems, as well as the previous work on the efficiency of computer programs in assisting groups, point to potential opportunities in using automated and conversational systems to facilitate group tasks. In this chapter, I present a framework for an automated group decision-making facilitation system guided by an embodied conversational agent that provides multiple types of facilitation during the group meeting lifecycle (opening, middle stage and closing [52]). This framework particularly focuses on group *decision-making meetings*, that happen synchronously and in the same place, as a ubiquitous example of group meetings.

I propose to develop an Automated Group Facilitation (AGF) system to support and facilitate group decision-making tasks. The AGF system is guided by a Virtual Facilitator (VF) that is an embodied conversational agent designed to deliver the services and interventions provided by the AGF system. The AGF system's goal is to address the common meeting problems identified by previous research on group decision-making such as: off-topic discussion (lack of structure), domination and unequal participation, conflict and tension, ineffective progress [188, 196]. In the following paragraphs, I review the essential processes required to support an automated group facilitation system.

3.1 Essential Processes to Support Automated Group Facilitation

3.1.1 Perception of the Virtual Facilitator as a Group Member

To act as a peer group member and to interact with other members of the group, the Virtual Facilitator (VF) should be able to simulate some human-like social behaviors. A virtual agent that is capable of conducting a quasi-natural conversation can be perceived as a social entity by the group members. Moreover, an anthropomorphic conversational agent with human-like face and nonverbal behaviors such as gaze and facial expression will allow participants to treat it as an intelligent social entity with a persona. For example, the agent can show emotions or active listening behaviors by moving her eyebrows, lips and nodding her head. More importantly, a conversational agent (CA) that takes the group facilitator role should be perceived as an authorized and intelligent member to make participants attend and adhere to her instructions. Justine Cassell suggested that the agent's embodiment helps users to better locate its intelligence and power [34]. Considering these requirements, I hypothesize that a CA with human-like embodiment would be more appropriate for a group facilitation role because it improves the perception of social behavior, intelligence and authority. I test this hypothesis in the first feasibility study.

3.1.2 Supporting Multi-Party Conversation

Another important feature of the group decision-making setting is the involvement of multiple humans in the interaction, so a VF should be able to support multi-party conversations. To communicate with multiple users, the agent needs to receive inputs from multiple sources, process them, detect the speaker and addressee, and understand the speech input to provide the right action or response. A virtual facilitator also requires turn-taking management capabilities. For example, she should know which participant is speaking, and show that she is listening. A VF should indicate her intention to talk and gazes toward each participant when s/he is expected to talk.

To satisfy the requirements of the automated group facilitation system, I propose a 3-layer framework to support all the group interactions and provide proper intervention (by the VF). This framework adapts the Smart Meeting Rooms' architecture[202] and integrates it to the conversational agent structure.

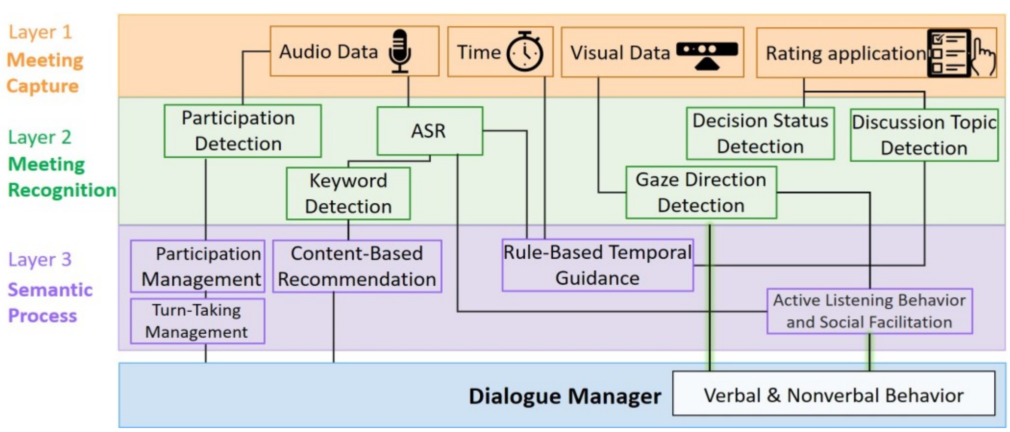


Figure 2. The AGF System Architecture for Decision-Making Meetings

3.2 Framework and Intervention Processes

Figure 2 shows the framework of my proposed Automated Group Facilitation (AGF) system. The AGF system obtains the audio data from microphones attached to each participant, to capture their speech in the meeting (**Layer 1** of AGF architecture). The system uses an in depth camera to see the group members and detect their attention direction. Participants are also provided with a simple application running on a two touch screen tablets in the room. The application provides support for the specific group task (e.g., tanking the items and sharing the decision status). The inputs from this interface allow the system to detect the stage of the meeting and provide some contextual data about the ongoing decision. In order to have pro-active, and just-in-time interaction with the group, the system relies on the second and third layers of AGF architecture: meeting recognition and semantic process. In the second layer (**Layer 2** of AGF), the collected data are processed and generate low-level information such as each speaker participation (from a voice activity detection module), keyword spotting, and conflict moments. In this second layer, the system also gathers data needed for agent's turn taking and nonverbal behavior (e.g., gaze). The third layer gets the analyzed information from the second layer and provides the intervention actions for managing structure, and conflict (**Layer 3** of AGF). For example, the 'Participation Management' module receives the participation data form layer 2 and passes a signal to the dialogue manager to nudge the less active participant at certain points during the meeting (e.g., at the end of each stage). Finally, the output of the third layer goes to the dialogue manager. For more details on each of the modules of AGF framework, please see Chapter 5.

In the group decision-making facilitation system, the CA provides active facilitation throughout the stages of the meeting. The agent's facilitation tasks are categorized into two groups: *General* meeting facilitation and *Decision-Making* facilitation. Table 1 shows how the agent delivers these two types of facilitation throughout each stage of the group meeting.

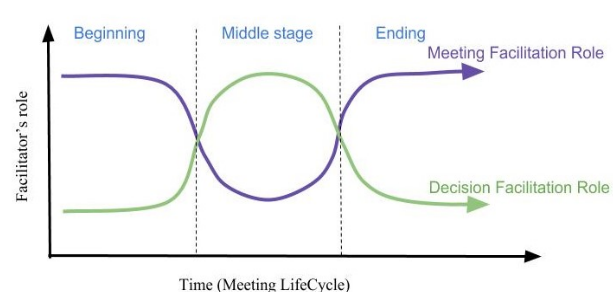


Figure 3. Facilitator's role in the group life cycle based on [71]

Meeting initiation: The agent initiates the session with greeting and introduction. She then introduces the parties and socializes with them (icebreaking). Next, she sets an agenda, ground-rules and norms, and defines the task, goal and decision-making process (grounding).

Middle stage: The agent attempts to enforce the structure of the decision-making process, to push the group to move forward and stay on track. She offers effective handling of possible disagreements, floor management, and engagement of less active members by encouraging equal participation. The system also provides computer-mediated tools for idea sharing and voting.

Emergence of ideas and conclusion: The agent keeps a record of group progress, summarizes the decisions, and concludes by announcing the final decision. Finally, she wraps up and generates post-meeting reports.

Table 1. shows all the services provided by the system in different stages of the meeting.

Assistance Type→	General Meeting Facilitation	Decision Making Facilitation	
↓Meeting Phase		Decision Process	Conflict Management
Beginning and intro	Greeting, making introductions, social chat (ice breaking), grounding, setting up goals and rules, describing agenda		
Maintaining (Middle Stage)	Presenting and following agenda, enforcing structure, balancing and engaging equal participation, recording progress	Providing information support, communication support, and rule-based temporal intervention	Detecting conflict, involving participants and resources, and managing conflict
Ending and wrap-up	Announcing conclusions and action items, thanking participants and making farewells	Providing a step-by-step report, summarizing, and wrapping up	

Douglas[59] described a variable role for the central person in a meeting depending on the group's development phase. In this design I follow Douglas's description for the facilitator's role, thus although the agent plays an important role in the beginning of the meeting and in setting the agenda, her general facilitation role is depreciated in the middle stages of the group's lifecycle when the group is more focused on the decision-making process. Instead, the agent provides more support for the decision-making process in that stage [Figure 3].

Last but not least, in this design, the group follows a 4-step decision-making process adapted from the 'Nominal group technique (NGT)[2]' in a small group.

1- Silent generation of ideas, reviewing the options and initial rating

2- Sharing idea and brief explanation

3- Group discussion: I) Criteria discussion, II) Eliminating unfavored options, III) Reaching an agreement

4- Final voting and emerging the options

In this research, I focus on two primary services of the automated facilitation system that are provided to improve the group decision making experience:

A. Meeting facilitation (Enforcing meeting structure and participation management)
B. Decision-making facilitation (Conflict management)

In the following paragraphs I describe how AGF framework supports each of these two services:

A) Meeting Facilitation

Previous research on meeting dynamics showed that having an agenda [177] and following it [164] are very essential for meeting efficiency. Also, one of the main tasks of group facilitators is to manage the meeting by keeping the discussions on track and ensuring the group follows the agenda [143, 188]. In AGF architecture (Figure 2), the Virtual Facilitator (VF) uses audio and inputs from the tablet application to track the meeting stage [42], and provides proper interventions based on the meeting stage, time and content. For example, to circumvent off-topic or inefficient discussion as main derailing factors in the meetings, the VF enforces the meeting agenda by asking targeted questions or reminding participants of the time. The system also analyzes the Automated Speech

---

[2] NGT is a procedural group process that is widely used for effective decision making. NGT process involves 5 stages to identify the problem, generation and sharing of solutions and making a decision.

Recognition (ASR) output and uses an off-the-shelf keyword detection module to provide content-based interventions.

Another common problem in meetings that can be improved by a facilitator is lack of participation or unequal participation [164, 188]. Prior work in CSCW community has also investigated how technology can enhance group collaboration [110]. In order to equalize the members' participation in AGF system, I use a Voice Activity Detection (VAD) algorithm to monitor each member's speaking time. The VF uses the output of this algorithm and engages the less active participants (e.g., by asking his/her ideas) and/or notifies the dominating members implicitly.

B) Decision-Making Facilitation (Conflict Management Intervention)

Group conflict is inevitable in collaborative decision making tasks, and it impacts the group feelings and performance [76]. There are different types of intragroup conflict that I reviewed in the related work sections. In this work, I focus on *task conflict* as the most prevalent type of conflict in newly-formed groups[200]. Task conflict results from disagreement over ideas and perspectives about the group task, and research has shown that unmanaged task conflict can impair group performance seriously [198]. The two main team development theories also [73] [185] confirm that conflict management in the middle stage is crucial for team performance.

For efficient conflict management, the AGF systems first needs an adequate model for detecting conflict moments. Prior research has investigated the use of verbal and non-verbal cues for detecting points of disagreement or conflict [23]. For example, Germesin and Wilson could detect agreements in a multiparty conversation with good accuracy by training a supervised machine-learning model on different voice features. In this work, I use deep learning-empowered methods to detect logical conflicts during the discussion. In order to detect the points of argument from the speech and voice data, I train a deep neural model on a public group conversation dataset. I model the data with different machine learning networks (e.g. LSTM and BERT) to examine the best model for detection of (dis)agreement from the verbal communication.

Upon detecting a conflict, the system attempts to manage it by providing a proper intervention. According to conflict management tips [158], the facilitator or the leader of the meeting should first assure that everybody's idea is heard. Then the virtual facilitator attempts to raise a wide range of alternatives and create a cooperative atmosphere for discussing different aspects of a topic (see chapter 7 for more details). In the Future Work section, I discuss other potential interventions (e.g.

offering private messages to each member to justify contrasting points of view.) that could be delivered by the AGF system for smooth handling of a conflict in a team.

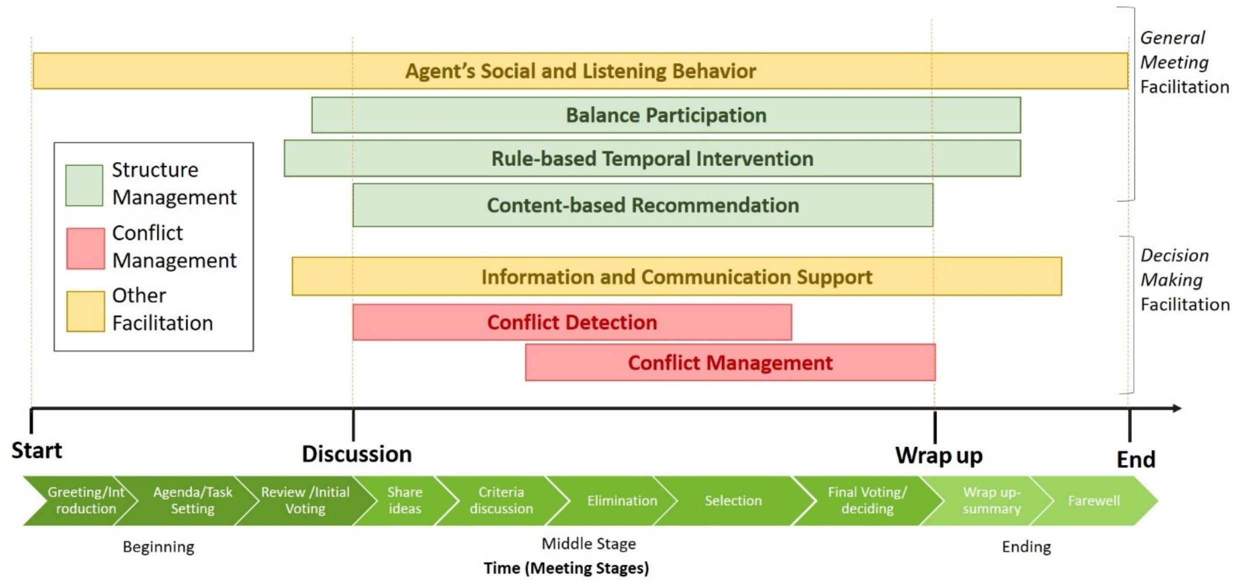


Figure 4. Overview of the Group Decision Making Facilitation System

Figure 4 shows an overview of the experimental procedure, showing how the system aids the group during the different phases of the meeting lifecycle.

Relation to Existing Work on Group Decision Facilitation

DeSanctis et al. defined three types of Group Decision Support System (GDSS) within the information-exchange perspective of GDSSs[54]. They suggested three possible support features for a GDSS when the task purpose is "CHOOSE" and the task type is "Preference". For such a task, a type1 GDSS may have a voting and ranking schema to remove communication barriers. A type2 GDSS can enforce a structured decision-making technique such as the Delphi method or the Nominal Group Technique to reduce the decision uncertainty and noise. Finally, a type3 GDSS for a preference task can apply rules and patterns (expert advice) to improve the decision quality. These rules include emphasizing equal time for sharing ideas, and/or limiting the time for each stage. Following DeSanctis's taxonomy, my AGF system provides two of the common services of computer-mediated GDSSs: information support and communication support. It also provides the voting schema, as well as guidance to keep the meeting structure and temporal rules.

Chapter 4: Prototype I – EXPLORING THE EMBODIMENT OF A GROUP FACILITATION AGENT

I am interested in increasing the ability of groups to collaborate efficiently by leveraging new advances in AI and Conversational Agent (CA) technology. Given the longstanding debate on the necessity of embodiment for CAs, bringing them to groups requires answering the questions of whether and how providing a CA with a face affects its interaction with the humans in a group. I explored these questions by comparing group decision-making sessions facilitated by an embodied agent, versus a voice-only agent. Results of an experiment with 20 user groups revealed that while the embodiment improved various aspects of group's social perception of the agent (e.g., rapport, trust, intelligence and power), its impact on the group-decision process and outcome was nuanced. Drawing on both quantitative and qualitative findings, I discuss the pros and cons of embodiment, argue that the value of having a face depends on the types of assistance the agent provides, and lay out directions for future research.

4.1 Introduction

Computer technologies have changed collaborative work in profound ways. Recent advances in Artificial Intelligence (AI) and Conversational Agents (CAs) spark new excitement for bringing technologies endowed with human roles and human- like behaviors into collaborative processes. By enabling inter- actions in a more natural form–conversations–CAs can potentially dissolve human-human and human-machine interaction boundaries by sensing, listening to and taking active roles in group activities. One of the roles a CA may take is to act as a group facilitator. Even if CAs cannot behave realistically like a human facilitator, many key functions of group facilitation (e.g., multi-party conversation monitoring [124], agenda setting [128], and preference elicitation [41]) are the targets of current research and technology development. I think the time is ripe to consider the key design issues for group facilitation agents. In this chapter, I revisit a fundamental question that the HCI and CA communities have asked before: "*Does a conversational agent need a face?*"

Many pioneers in the research communities are advocates of embodiment. Some of the pro-arguments were documented in the 2000 book *Embodied Conversational Agents* edited by Justine

Cassell et al. [34], where she argued that having multiple modalities including gaze, face and gesture is the only way to attain human-like intelligence. Otherwise users would have trouble *locating* both task-related capabilities and social intelligence, because we are wired to exhibit social behavior such as turn-taking and affect. However, the CAs entering the mainstream market in recent years–the most popular being Apple's Siri, Amazon's Alexa, Microsoft's Cortana, and IBM's Watson–do not have embodiment beyond simple icons or inanimate objects.

Many argue that, even putting costs aside, embodiment may not be necessary for CAs [50, 83, 139]. Empirical evidence about the necessity of the embodiment is mixed. Some literature [15, 75, 176] suggests that embodiment could improve subjective impression of the agents such as trustworthiness, and thus interaction engagement, but not necessarily objective performance of the tasks that the agent assists in [52]. There seems to be a pattern that for tasks that require continuous engagement (e.g., tutoring), embodiment of the agent improves task performance [8, 137]; whereas for more "sporadic" interactions where agents only occasionally respond or prompt, embodiment may not benefit task effectiveness and it is not necessary to have an agent continuously "being there" [50, 83]. These arguments may underly the design of popular text- or voice-only personal assistant type of CAs.

What should we expect for a group facilitation agent? On the one hand, if the main function of a facilitation agent is to enforce a structured and balanced process, the system would continuously sense the context, but only interact sporadically. In this case, embodiment may not be necessary. On the other hand, if embodiment leads to more positive social perception of the agent, the benefit of embodiment may go beyond that observed in individual interaction settings. First of all, because enforcing structure implies attention and compliance, a more socially favorable agent may be more effective in im- proving group processes and outcomes. Secondly, affective benefit is often emphasized in technologies supporting collaborative work because it can improve social and collaborative process [130], and positive affect brought by CAs has been observed in casual social settings [159]. Furthermore, I postulate that embodiment may create a stronger sense of presence as in continuously "being there", especially in a group set- ting. Perceiving an additional entity "being there" may impact both group interaction and users' interaction with the agent. Whether such impact is positive or negative for collaborative tasks invites empirical inquiry.

These questions motivated my study to explore impacts of the agent's embodiment on group facilitation tasks. I designed and developed an embodied conversational agent that facilitates a group decision task by enforcing a structured discussion process. I conducted a between-subject, Wizard-of-Oz style experiment with 20 user groups, in which half of the groups experienced the embodied agent, and the other half experienced a voice-only version of the agent. Through survey responses, I first examined whether the positive effect of embodiment on social perceptions of CAs in individual user settings still holds in a group setting, and then studied how the embodiment impacted the collaborative task. I complemented my quantitative results with qualitative focus group interview data. My work was guided by the following research questions:

RQ1: How do different designs of the agent's embodiment (voice vs. embodied) influence subjective social perceptions of the agent (rapport, trustworthiness, intelligence and power) in a group setting?

RQ2: How do different designs of the agent's embodiment (voice vs. embodied) impact: **a)** the group decision outcome, **b)** participants' interaction with other group members, and **c)** participants' interaction with the agent in a group setting?

In the remainder of the chapter, I will first review related work that motivated my study. I then present how I designed the group facilitation agent, the Wizard of Oz experiment, as well as the experiment methodology. The result section starts with examining participants' subjective perceptions from survey responses to answer **RQ1**; then I report analyses on the process and outcomes of group decision making activities related to **RQ2**. This work contributes to the field by providing a comprehensive account of the impact of CA's embodiment in a group collaboration setting, and design considerations for conversational agents that go beyond supporting individual interactions to supporting group collaborations.

4.2 Related Work

This study is mainly informed by two sets of literature: the work on designing computer-supported systems to facilitate group decision-making tasks (that I reviewed in chapter 2), and the work on conversational agent design, in particular, the design of agents' embodiment and its impact on user perceptions and behaviors.

4.2.1 Conversational Agents and Embodiment

Conversational agents (CAs) are computer systems designed to interact with users through natural conversations. Numerous researchers have discussed the design, development, and evaluation of CAs that are designed for : 1) Task-oriented personal assistance services via short-term interactions such as question-and-answer services; or 2) Goal-oriented interactions to provide training, counseling content, or interventions to help users achieve goals. I consider a group facilitation agent to belong to the latter category.

Goal-oriented CAs are used in various domains such as health care [17] and education [183]. For example, Graesser et al. developed AutoTutor, an animated pedagogical agent [74] and demonstrated that the agent significantly improved learning outcomes by engaging students in an interactive conversation [183]. Although the research on CAs focused mostly on individual interactions, some HCI researchers explored CAs interacting with multiple users in group settings. For example, Kumar and Rose proposed a new architecture for building ped- agogical agents to enable them to handle complex interaction dynamics in a multi-learner collaborative environment [115]. Bohus and Horvitz also proposed a model to represent a CA's turn-taking behaviors by gaze, gestures and speech in multi-party conversations [18].

Although text-based un-embodied CAs have a longer history, the HCI community has largely focused on studying embodied conversational agents (ECA). One of the most prominent advocates for embodiment is Justine Cassell. She argued that in human communications, the body "*embodies intelligence*", both to serve propositional goals—conveying information, and to serve interaction goals—regulating the communication process. For computer agents, the multiple modalities of embodiment (e.g., verbal, gaze, gesture) can not only provide better means to manifest social intelligence (e.g., trustworthiness and rapport), but also help users locate the domain specific intelligence and capabilities [32, 33]. Therefore, ECA can serve as an interface for more intuitive and engaging interactions [19]. To illustrate, Cassell et al. developed REA, an ECA with a model to recognize and generate verbal and non-verbal behaviors, with which REA can manage turn taking, provide feedback, and repair conversations. Another argument for embodiment is the "persona effect" proposed by Lester et al. [119]. It suggests that "personification" with human-like behaviors may improve users' various perceptions (e.g., trustworthiness) of tutoring agents, further help with users' engagement, and ultimately improve users' learning outcomes.

However, empirical studies provided mixed results about the impacts of CA's embodiment. On the positive side, many studies showed that visual images and embodiment of CAs significantly improve users' perception of social presence [147], motivation [11], entertainment [112] and trust [15]. Embodiment has practical benefits as well. For example Walker et al. showed that users who interacted with a talking face spent more time, made fewer mistake and wrote more comments, compared to those who had text-only displays [192]. Given that education is a main area of focus for CAs, many studies examined animated pedagogical agents and found them to improve students' learning outcomes [136, 137].

On the negative side, Hasegawa et al. studied a direction- giving service provided by an ECA and a GPS system,[50] and found no difference in the user performance. Similarly, Hauslschmid et al. compared two designs of the control panel in an autonomous driving car, an avatar versus a regular visual display, and found no differences in user trust or other user experiences [83]. In examining how substantial the persona effect is, Mulken et al concluded that there is a difference between subjective measure (e.g., agent credibility and perception of the experience) and objective measures (e.g., comprehension and recall). While an animated agent showed positive effect on subjective measures, no effect on the objective measures was found. This conclusion was echoed in a 2007 meta analysis on 46 studies, revealing that embodiment has significantly larger effect size on the subjective impression of the agent than behavioral responses [62].

Reviewing the complex results regarding the effect of agent embodiment points to a divide between subjective perceptions and objective behavioral responses, and a divide between different application contexts. In contexts that require a user to be continuously engaged with the agent such as tutoring, the benefit of embodiment is more evident. While for sporadic interactions such as receiving instructions from the agent, the benefit of embodiment is less clear. The group facilitation agent seems to be more aligned with the latter. However, in presenting itself as a human-like facilitator, many factors in the subjective perceptions may come into play. For example, if perceived to be more socially connected, an agent may better engage users. If perceived to be more trustworthy or powerful, it may work more effectively to enforce structures. The group context may also result in subject perceptions and behavioral responses not observable in individual contexts. In this work, I set out to explore the impact of various aspects of an agent's embodiment in a group context and to understand the interplay between them. I aim to contribute both to practical guidelines for designing CAs in collaborative contexts, and new knowledge on the effects of embodiment.

4.3 Research Method

I developed a conversational agent and designed a Wizard- of-Oz experiment[3] [49] to explore how different embodiments of the agent (voice-only vs. embodied) impact users' perceptions, collaboration process and outcomes in a group decision making task. In the following subsections, I first describe the experiment task and design of the CA. I then present the facilitation protocol of the agent. Lastly, I lay out the details of the experiment design.

4.3.1 Experiment Task

I chose to simulate a hiring decision task as it is a common scenario for group decisions in the workplace, and previous work used similar tasks to study group decision processes (e.g., [61]). For the task, a group of two participants were asked to select the best candidate from a set of five resumes for a user experience design internship position. I created fictional resumes based on the real resumes of applicants for that position. Participants had 30 minutes to review, discuss and select the best candidate. The agent served as the facilitator of the decision-making session. The participants interacted with the agent through natural conversations.

4.3.2 Embodied Conversational Agent Design

I designed an embodied conversational agent (ECA) named CASSY (Collaborative Agent for Decision Support System) to facilitate group decision-making sessions. I used a commercial avatar toolkit to develop CASSY, a 3D avatar with a humanoid face (cut off above the chest, focusing on the facial area) and synthetic text-to-speech voice. The agent is designed with the look of a young female professional. She has an animated face with a range of non-verbal behaviors such as directional gaze, eyebrow raise and head-nod, with which she can display a variety of facial expressions such as happy, and apologetic. CASSY is projected on a Beam Pro system-A professional telepresence robot[4], both to fit a face- area representation and create a sense of mobility to enhance her presence (Figure 5).

[3] A Wizard of Oz experiment is a research experiment in which users interact with a computer system that they believe to be autonomous, but the system is actually being operated or partially operated by an unseen human being.
[4] (Manufactured by Suitable Technologies Beam Pro: http:// suitabletech.com)

At the present time, conversational technologies in the simplest form consist of four main components: 1) Speech recognition to convert speech input to text; 2) Dialogue understanding component that maps the raw text to an *intent* known to the system (when a user says "hello", the system understands it means "greeting"). 3) Response generation component that generates a response based on the understood intent (user's "greeting" is mapped to the agent response of "hi, how are you?"), and 4) Text to speech module that speaks out the response. Embodied conversational agents have an additional module to generate the non-verbal behaviors (e.g., BEAT [35]).

In this study, I adopted a Wizard-of-Oz approach, where a human wizard controlled CASSY. The human wizard aimed to replace the first two components: speech recognition and intent understanding. Simply put, when the human wizard heard a participant saying "hello" (speech recognition), s/he would send a command to CASSY selecting the intent "greeting". Then CASSY's automated components took over, generating the response and speaking it out. Wizard-of-oz is a commonly used approach to study user interactions with conversational agents [15, 26]. Not only does it reduce development cost to make design choices, but also it is often necessary for experimental studies to control for noises from system performance variations. State-of-the-art speech technologies still fail to give perfect performance in real-world settings, especially in a group context where multiple parties may talk simultaneously. Moreover, voice recognition errors can vary largely between individuals due to different talking style, accent, the volume, and these can be further amplified in a group setting.

With the Wizard-of-Oz setup, the design of CASSY focused on two parts. One was to design the functional knowledge of a facilitation agent. That is, what intents are available for the wizard to choose from, both intents from understanding a user utterance to the agent (e.g., "greeting") and intents inferred from monitoring ongoing group discussion (e.g., "debating on candidate 1"). The other was to design the agent's response once an intent is selected by the wizard, both in terms of the conversational content and non-verbal behaviors. The Watson Conversation API was used to build the response generation part. The text-to-speech and non-verbal behavior modules are part of the agent toolkit. I used a constrained Wizard-of-Oz protocol [161] with a small set of available intents (see next section), so that fully automatic systems can be built based on the protocol, and more importantly, the agent would exhibit a realistic level of intelligence to study user responses.

If users asked questions beyond the protocol, the wizard either answered "sorry I don't understand," or ignored the question (to simulate the behavior of a potential automated system).

Figure 5. Experiment setting: CASSY projected on a Beam telepresence robot, and facilitating a group decision making task with two participants. In front of the participants, there are job description, rating sheet, and resumes. Two iPads on the table se

Facilitation Functionalities and the Wizard of Oz Protocol A facilitator's main goal is to facilitate shared understanding and eventually consensus [29]. Group decision literature recommends an effective process to involve: 1) clear under- standing of the problem, 2) full exploration of each alternative, and 3) comparison of alternatives [63]. I designed the agent facilitation protocols with these objectives in mind, and also used elimination techniques (a common strategy in decision making) to winnow out candidates. In general, CASSY pro- vides three types of support:

Decision Making Facilitation: The agent enforces a structure of the decision process by guiding the group through pre-defined six steps. Example actions include: tracking decision states, suggesting actions and opinion exchange, summarizing decisions.

Meeting Facilitation: The agent engages in time management and participation management to improve meeting efficiency. Example include: Ice breaking, managing time, agenda description, turn taking management.

Social Interaction: The agent exhibits active listening behaviors to express interest and social support. Examples include: greetings, short verbal phrases to reflect under- standing and support (e.g., “I see”, “I agree”), and facial expression, (e.g., gaze at the participants while they are talking, nod with a smile, or confusion).

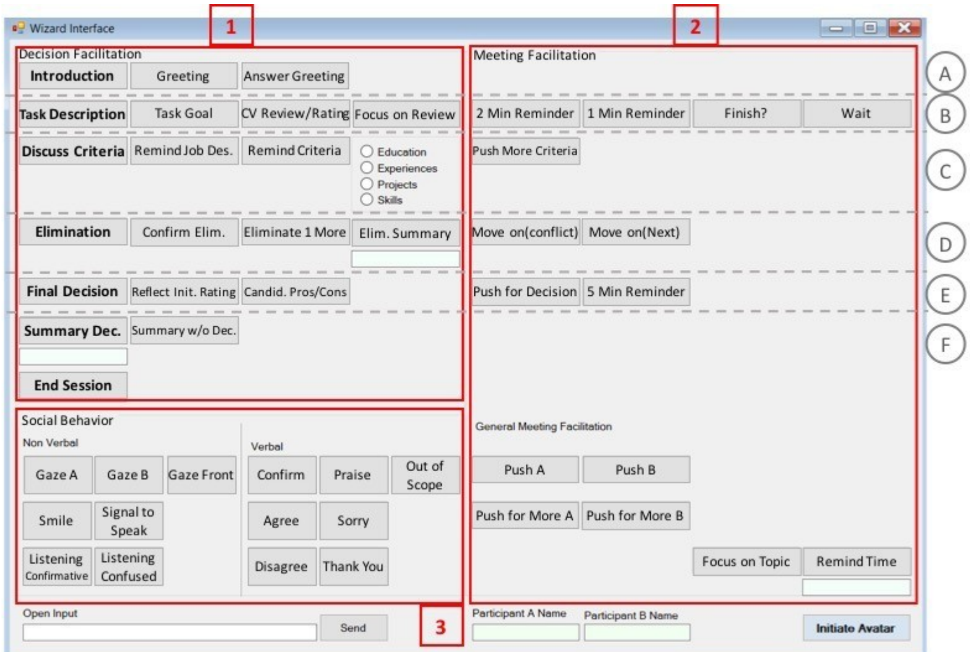


Figure 6. The wizard interface. 1-3 shows the three forms of agent’s services, and A-F shows the six states of each decision-making session.

Below I present a procedural view of the protocol, which covers most of the decision facilitation and meeting facilitation the wizard provides. I describe how the wizard selects intents on the Wizard interface (Figure2). The first author acted as the wizard in all experiments while sitting in a separate room, listening and watching the conversations. The wizard practiced the protocol in multiple pilot studies, which were reviewed by the whole research team to ensure consistency.

Introduction: CASSY initiates the conversation by introducing herself, inviting the participants to introduce themselves, and greeting them with their names. (Figure6-A).

Agenda Description and Resume Review: CASSY describes the agenda, and then asks the group to review the resumes independently for five minutes and to rate them on a rating- sheet. She reminds them of the time at 3- and 4-minute marks. At the 5-min mark, she confirms their readiness to proceed. On occasions when participants request more time, she agrees to extend it for an additional minute (Figure6-B).

Criteria Discussion: CASSY suggests that the group discuss the hiring criteria. After the initial exchange, she prompts them to consider the job description, if not dis- cussed. Once the group has discussed the four main criteria on the resume (e.g., education, and skills) she moves to the next step; otherwise, she prompts them to consider the missing criteria. (Figure6-C)

Elimination of Unfavored Candidates: CASSY invites the group to go through the candidates and eliminate the unqualified ones. When participants converge on an elimination, CASSY confirms that decision. If participants cannot decide about a candidate after two minutes, CASSY suggests that the group discuss the next candidate. After all candidates are discussed, if fewer than two candidates are eliminated, CASSY suggests the group to eliminate one more. Finally, CASSY summarizes the remaining candidates, and moves on to the next step. (Figure6-D)

Decision on the Final Candidate: CASSY asks participants to select the best candidate. If no consensus is reached, she suggests the group reflect on their initial voting, or discuss pros and cons of the remaining candidates. When either 30-minutes is up, or the group reaches a consensus, CASSY moves to the last step (Figure6-E).

Exit the Experiment: CASSY summarizes the session, either with a final decision or without a consensus, and thanks the group for participating. (Figure6-F).

4.3.3 Experimental design and procedure

I adopted a between-subjects design to compare group decision making sessions facilitated by an embodied agent (avatar condition), versus a voice-only agent (voice condition). Each experiment session was randomly assigned to one of the two conditions. In the avatar condition, CASSY was projected on a Beam telepresence robot and the voice was played out from a bluetooth speaker on top of the Beam (Figure1); in the voice condition (no Beam), CASSY's voice was played via the same speaker on the table. Figure1shows the setup of the experiment. Two folders containing the hiring advertisement and resumes were placed on the table before the participants came in. Prior to the study session, consent was obtained from the participants. Post-study questionnaires and a focus group interview were collected after each session. Two iPads (with covered screens) were used to record the the participants video and audio during each session. At the end, the participants were given a $12 compensation as well as a debriefing document disclosing the Wizard-of-Oz design. In total the session and interview took between 45 and 60 minutes.

Post-experiment Focus Group

To gather more feedback and insights on their experience, I conducted 20-30 minute semi-structured focus groups with my participants. I started by asking questions about their overall impression about the decision process and the agent's facilitation. I also inquired about their reactions for different types of agent actions (e.g., meeting facilitation, social behaviors). Lastly, participants were asked about how they would desire the agent to be improved, and new features they would like to add to the system. Interviews were audio recorded.

Participants

40 participants were recruited from an IT enterprise (IBM headquarter in Yorktown, NY) by posting advertisements in the campus and in online forums. The participants consisted a mix of full-time employees and interns who were college or graduate school students (35%). Participants were randomly matched together to form 2-member groups. 60% of the participants were male, 20% between 18-24 years old, 57% were between 25-34 year old and the rest were older than 35. To minimize the effect of gender and expertise on the group performance, I ensured that the ratio of mixed- gender/same-gender groups, and the ratio of employee/student groups were equal in the two conditions (Fisher's exact test

$p = 1$). I avoided grouping people from the same department, or different expertise levels together. 90% of participants

did not know each other before the sessions.

Survey Measures

My measurements consist of subjective responses from a survey, and objective measures reflecting the decision process and outcome. In this section I present the survey measures. Details of the objective measures will be introduced together with the results. Guided by previous work on user perceptions of agent systems, I asked participants to rate the agent's rapport, power, anthropomorphism, intelligence, and trust. To complement the objective measures on group processes, I asked participants to report subjective evaluation of their experience with the decision making process, and towards their collaborator. I also collected demographic information such as gender and race at the end of the survey.

User Perceptions of the Agent

To answer RQ1, I assessed users' perception of rapport with the agent, which is considered a key dimension for "socially aware agents" in recent work [47,64,23]. Rapport refers to a feeling of connection and bonding in an engaging interaction [23,64]. Based on an instrument to measure human-agent rapport from previous work [43], I asked participants to indicate how much they agree or disagree with five items (α =.84). Sample items include: "The agent seemed engaged in our discussion", "I felt I had a connection with the agent", and "I felt the agent was NOT paying attention to what I said".

I used a list of semantical differential scales to measure a number of other dimensions regarding the social perception of the agent, including intelligence, anthropomorphism, and trustworthiness. I adapted validated scales on these dimension presented by [2] and [46] by selecting 2 items for each scale. I asked participants to rate the agent on pairs of antonyms criteria such as *ignorant/knowledgeable* and *unintelligent/intelligent* for intelligence, *human-like/machine-like* and *conscious/unconscious* for anthropomorphism, and *unreliable/reliable*, *untrustworthy/trustworthy* for trustworthiness. Given the task of meeting facilitator as enforcing structure, I added power as an additional dimension, measured by *weak/powerful*, and *lacking confident/confident*. All the above agent perception ratings were based on 7-point Likert scales (1= strongly disagree, 7= strongly agree). I calculated the average ratings of items for each perception dimension to be the scales used in the analysis.

Decision Making and Group Interaction

To answer RQ2, I asked participants to rate their preference for each candidate before and after the session on a 7-point Likert scale, and studied the rating changes as decision outcome measures. I also asked participants' subjective evaluation of the group decision performance in the survey. As suggested by [24] I asked them to indicate to what level they agree with four statements about the decision (e.g., *"I found the decision-making task to be difficult"*, and *"we reached a decision efficiently"*). Inspired by [52] I asked participants to report their perception of their partner with statements such as; *"I found it pleasant interacting with my partner"*, and *"I found my partner and I shared many similarities"*.

Statistical Analysis

The main part of my quantitative analysis is to compare participants' ratings of agent perceptions between the two conditions—avatar vs. voice-only. Participants experienced the agent in groups of two, so ratings from each pair might be affected by their common experience (e.g., a group made a smooth decision versus a difficult one). To account for such group effects, I utilized Linear Mixed Models (LMM) by having condition as a fixed factor, and group as a random factor to control for their associated intraclass correlation (i.e., random intercept models [48]. Random slope was not considered as group is nested within the conditions). I used unstructured covariance matrix for random effects. Although debates exist, there is consensus that parametric tests such as regression are generally robust with Likert scales [42], especially for the 7-point composite scales I used. I also tested the data to ensure the residuals meet the assumptions of linear regressions. Common non-parametric tests for Likert scales such as Mann-Whitney U test cannot account for the random effect of group, and my exploratory analysis showed they yielded the same conclusions as LMMs. In the rest of the chapter, I report results from LMM, whenever the random group effect needs to be considered. I ran LMMs with the *lme4* package of R, with the default restricted maximum likelihood estimation. To report p-values of variable effects, I used likelihood ratio tests (R procedures as seen in [61]), a common approach for significance testing of LMMs

4.4 Results

In the following sections, I first discuss how the embodiment influenced social perceptions of the agent compared to a voice-only condition (RQ1), and then examine the differences in decision outcomes and group interactions between the two conditions (RQ2). Last but not least, I present my qualitative findings and further discuss the pros and cons of agent's embodiment in the group context (RQ1 and RQ2).

4.4.1 Agent perceptions (RQ1)

To answer RQ1, I examined participants' subjective perceptions of CASSY based on their survey responses. Specifically, I compared the ratings of the scales of rapport, anthropomorphism, intelligence, power and trust between the avatar and voice-only conditions, measured as discussed in the method- ology. For each scale, I conducted a linear mixed model regression by including the condition (voice vs. avatar) as a fixed factor, and group ID as a random factor. I removed one group that appeared to be an outlier with multiple agent perception scales (e.g. rapport, trust, and power). In Table2, I report on the statistics of the effects of embodiment with the five agent perception scales (analysis detailed discussed in the "Statistical Analysis").

The results showed that the positive effect of embodiment is statistically significant or marginally significant for all dimensions of social perceptions[5]. For RQ1, I can conclude that embodiment improved participants' subjective perceptions of the agent. Adding a face not only made the agent show more rapport and human-like characteristics, but the embodied agent was also perceived as more intelligent, trustworthy and powerful as a group facilitator.

4.4.2 Decision performance (RQ2-a)

Next I examine whether the embodiment and improved subjective perceptions had an impact on decision outcomes. The goal of the agent was to facilitate a consensus building process. Therefore, I used consensus and opinion shift as measurements of decision outcomes. Before and after the group discussions, participants were asked to independently rate the favorability of each candidate (from 1=not at all to 7=a lot). I examined two kinds of rating shift: 1) Consensus shift, for which I used the difference between the intraclass correlation coefficients (ICC) of a pair's ratings before and after the discussion. ICC is a statistic reflecting inter-rater agreement (i.e., consensus) with ordinal ratings, with a value ranging between -1 (no agreement at all) to 1 (complete agreement).

Table 2. Results of LMMs on agent perceptions. Statistical significance is obtained from likelihood ratio test (logLik, chi-square and p-value reported), where the fixed-effect is tested against a null model with random effect only (using maximum likelihood estimation), and random effect against a null model with fixed effect only. $p < .05$ is considered significant*, $< .1$ marginally significant†. Random effect is consistently included regardless of its significance based on the assumption of a group based study.

	Descriptive statistics		LMM statistics for fixed effect (condition)							Random effect (group)			
Agent Scale	Avatar M(SD)	Voice M(SD)	(intercept)	β	SE	CI (95%)	logLik	Chi-square	p-value	Variance	Std. Est. Dev	Chi-square	p-value
Rapport	5.57 (0.62)	4.64 (1.34)	3.63	0.93	0.39	[0.10,1.70]	-55.05	5.44	.02*	0.87	0.93	0.72	0.39
Trust	5.22 (0.84)	4.42 (1.07)	4.42	0.80	0.37	[0.01,1.50]	-51.10	4.57	.03*	0.33	0.58	1.69	0.19
Power	5.05 (0.85)	4.42 (1.09)	4.42	0.63	0.32	[-.06,1.32]	-52.40	3.71	.05*	0.05	0.22	0	1
Intelligence	4.97 (0.94)	4.20(1.43)	4.20	0.77	0.44	[-.14,1.69]	-60.61	3.22	.07†	0.27	0.52	0.29	0.59
Anthropomor.	4.22 (0.94)	3.55 (1.03)	3.55	0.67	0.39	[-.14,1.40]	-51.39	3.11	.08†	0.41	0.64	2.55	0.10†

[5] following common practice for small-scale lab experiments [16], I consider $p < 0.05$ to be significant, or < 0.10 marginally significant

2) Individual shift, for which I calculated the ICC of each participant's pre and post ratings. So a higher individual ICC indicated less individual opinion shift. I ran a T-test on the consensus shift, and a linear mixed model (group as the random factor) on individual shift to compare decision outcomes between the avatar group and the voice group.

I also looked at the groups' top choices. All groups but one reached agreement on the top candidate. While I did not intentionally design the study with a best choice, I identified candidate 2 to be the majority choice (55% groups) and examined percentage of groups selecting candidate 2 in each condition, and compared them with a chi-squared test. In addition, I compared the time taken to reach the decision in the two conditions (Voice: *Mean*(*SD*) = 23.0(3.43), Avatar: *Mean*(*SD*) = 21.7(4.76)), and individuals' self-reported confidence score shift. All statistics are shown in Table 3.

According to these results, I saw no significant improvement that embodiment made on reaching consensus, opinion shift, decision confidence or efficiency (time). Moreover, I saw no significant difference on self-reported decision satisfaction, for which I asked participants to rate on decision easiness, satisfaction, success and efficiency in the survey (averaging the four items, LMM: β = 0.16, *SE* = 0.44, p = 0.72). I conclude that the embodiment did not make significant impact on decision outcomes.

4.4.3 Group interactions (RQ2-b)

I further explored whether the embodiment had any effect on interactions between the group members. By looking into the survey items on perception of partners, I found two items that show improvement in the avatar group: "*I found my partner and I shared many similarities*" (β =0.99, *SE* = 0.465, $t(36)$ = 2.14, p = 0.05) and "*I made efforts to respond to my partner's questions and suggestions*" (β =0.62, *SE* = 0.37, $t(36)$ = 1.67, p = 0.10). These results led us to hypothesize that the improved social presence of embodiment had a positive effect on group interactions. A previous study also showed that the presence of embodied agents may create a sense of "being watched" [104]. In a group setting, the additional presence could possibly have created a sense of social pressure that increased individuals' tendency to actively contribute and share opinions. A manifestation of such a tendency could be more equal contribution between the pair, so both parties experienced a more amicable and engaging discussion.

Table 3. Results of decision outcome measures (TT:T-test, MM:mixed model linear regression. [4]$\beta = 0.04$, SE = 0.143 , [5]$\beta = -0.2$, SE = 0.289

Decision outcome	Avatar Mean (SD)	Voice Mean (SD)	Test	*P*-value
Consensus shift	0.44 (0.38)	0.56 (0.35)	TT	0.49
Individual shift	0.64 (0.34)	0.59 (0.40)	LMM[4]	0.76
Majority choice	%50	%60	$\chi 2$	1
Confidence	1.1 (0.79)	1.3 (0.98)	LMM[5]	0.49
improve Time	21.6 (4.72)	22.8 (3.55)	TT	0.53

To test this hypothesis, I compared the contribution equality of pairs between the two conditions. After transcribing all discussions, I counted the number of turns and the number of words from each participant through a text analytic process. I calculated contributing percentages of words and turns from the less talkative participant, divided by the words and turns from the more talkative partner (i.e., a higher percentage indicates more equal contributions). As hypothesized, by one tailed T-tests, I found equality of turn contribution ($t(18) = 1.79$, $p = 0.04$) and words contribution ($t(18) = 1.5238$, $p = 0.07$) to be higher in the avatar group than in the voice-only group. Together with the self-reported positivity on sharing similarities and making efforts in responding to the partner, the results suggested that the enhanced social presence of the embodied agent had a positive effect on the interaction process, despite a lack of effect on the decision outcomes.

4.4.4 Agent interactions (RQ2-c)

Lastly, I looked into participants' interactions with the agent. My initial examination of the transcripts revealed that participants were mostly compliant with the agent's commands, so the interactions were generally structured and showed no evident differences between the two conditions. One exception I observed is that the embodied agent seemed to encourage participants to engage in proactive interactions, which the current version of agent was not explicitly designed for.

To test this trend, I coded proactive interactions in the transcripts, and categorized them as three groups:1) *social interactions*, such as "*what do you do here?*" or "*thank you*", which typically happened at the beginning or end of the discussions;

2) *asking questions*, such as clarification like "*what do you want us to do?*" or "*which candidate?*"; 3) *requests*, mostly exemplified by asking to extend the time of resume reviewing. The Wizard-of-Oz protocol allowed only simple response such as "ok" or "candidate two", otherwise the agent

would ignore the proactive interactions. The statistics of proactive interactions in each condition are listed in Table 4. In total, proactive interactions more than doubled in the groups with embodied agents and one tailed t-test showed marginal significance ($t(18) = 1.6015$, $p = 0.06$). I conclude that the embodiment invited more proactive interactions from the participants.

Table 4. Average occurrences of proactive interactions with the agent; standard deviations are in the parentheses

Condition	Social	Question	Request	Total
Avatar	1.5 (1.71)	3.3 (3.65)	1.1 (1.29)	5.9 (5.61)
Voice	0.5 (0.85)	2.1 (1.60)	0.3 (0.48)	2.9 (1.91)

4.5 Qualitative Findings

The results above provide evidence that the embodiment improved users' subjective perceptions of the agent, and impacted group interactions and user interactions with the agent. In this section, I present qualitative findings that shed further light on the impact of embodiment. Moreover, by inquiring about the general user expectations for a group facilitation agent, I reflect on the implications of embodiment for the ap- plication domain. All focus-group interviews were transcribed and coded using thematic analysis techniques. The results of my qualitative analysis converged to three main themes related to the benefits users see in the embodiment of the agent, and three themes of contributions they expect a facilitation agent to bring to group decision processes.

What to expect from embodiment?

Embodiment for enhancing presence: Participants commented that embodiment enhanced the agent's presence, making it more of an active part of the group discussion instead of lurking in the background, as illustrated in this quote:

> *[P7]: "...it's more like a group discussion I feel like. [P8]: I agree. In a group discussion you hope that it does actually have a face indicating that now she is participating in this conversation rather than just an object on the table... [P7]: otherwise I feel like she just not existing here. I can just like talk, and (she is) in the background."*

This observation echoed a previous study showing that the mere presence of an embodied agent can create a sense of "being watched" [104]. I suggest this could be an explanation for the observation of more equalized contributions in groups with avatar, although the validation awaits future research. This stronger sense of presence may be especially important for group facilitation tasks where, unlike individual interaction with CAs, users' locus of attention is the human partner. The enhanced presence can improve engagement, while a lack of presence may weaken the influence of the agent.

However, not all participants favored a stronger presence of the agent and some preferred staying focused on the discussions and having the agent in the background. For example, P35 preferred the agent not to have any visual presence, because he wanted to *"interact with it as a machine, not a human"*. He added: *"it should be more an assistant rather than a part of the decision...Decisions should be made by us"*. P16 experienced a voice-only agent and commented: *"it is the right balance...it's better to stay in the background and only come out at a certain point, not interrupting too much"*. Such arguments can testify for the social impact of agent's embodiment and suggest that preference for enhanced presence may depend on the task.

Embodiment for locating social-interactional intelligence: One convincing argument for embodiment is that interactional functions regulating communication processes (e.g., turn-taking, interrupting) are best served in multiple modalities, so agents can leverage users' familiarity with non-verbal behaviors to improve the interactions [5]. This argument was frequently echoed in the interviews, for example P29 said:

> *"she was trying to make eye contact, moving eyebrows. Looking into your eyes... To me it was good. Like more engaging."*

Several groups also discussed how the agent's visual presence can be a useful modality to indicate her intention to talk. For example, P1 and P2 agreed that "*(a face) would help with knowing when it was about to talk*".

This may be beneficial for facilitation tasks, because by nature facilitation involves frequent floor-taking and interrupting. However, interestingly, some participants in the embodied condition commented that the additional modality became a source of confusion with the system latency. When they were unsure whether the agent was going to continue or they should take the floor, their attention was fixed on the face but were un- able to locate indicative cues. Interestingly, there is evidence in the survey response that participants found the agent less interruptive when she had a face (M= 3 , SD= 1.68) versus the voice-only agent (M= 3.95, SD= 1.82; p = 0.09), but they might be more annoyed by the latency in the embodied condition (M = 4.95, SD = 1.93, p = 0.16, not statistically significant (M = 5.75, SD = 1.61) compared to the voice condition but indicating a trend).

In the survey responses, I found that participants perceived a stronger sense of social connection (rapport) with the embodied agent. Qualitative results suggest that one reason could be that participants were able to locate attention and emotions from the additional modalities of gaze and facial expression. P29 said: "*The good thing is that this is one of the few agents that I've seen having emotions. So that's really a plus!*"

Embodiment for exhibiting task capabilities: Survey responses showed that participants perceived the embodied agent to be more intelligent and trustworthy, suggesting that embodiment might also improve perceived capabilities. Although the agent currently provides little informational support, I found evidence that participants perceived better *understanding* and more *knowledge* of the embodied agent. For example, group 2 said they *"prefer a realistic face"* because they feel *"she understands us because she looks like a human"*, and *"it raises my confidence on her."*

This improvement in perceived intelligence together with the enhanced presence may explain why embodiment invited proactive interactions. For example, P37 (voice condition) mentioned that he might take a agent's human-like embodiment as a signal that she is capable of answering questions. He said: "*I don't think (voice-only) works very well. I think a depiction of a real human ... would be much better. Because it's capable of [answering] the questions.*"

What to expect from a group facilitation agent?

Structure management: As participants were asked to reflect on the values of a group facilitation agent, many con- firmed needs for CASSY's targeted function—managing structure. Participants described their experiences in group meetings where attendees *"don't follow agenda"* (P7) and *"are hijacked"* (P38) by a few dominant people. They also appreciated CASSY's functions for keeping time, keeping the discussions on track, and encouraging everyone to express his/her ideas by providing step-by-step instructions. For example, group 4 (avatar) compared the meeting facilitated by CASSY to their prior group meeting experiences:

"The decision-making is much faster than what I really have in the real life, most of the times I cannot reach an agreement and then I go past...people diverse and digress...but the agent is clever and just coming, cut the conversation and force you to move to the next phase."

Several groups also brought up how the agent could be helpful in balancing the discussion without creating tensions or impairing group dynamics. For example, P38 said:

> *"If my manager is going on and on from the topic and hold up the meeting, I don't want to be the guy to stand up and say 'Can we get back to what we're supposed to be talking about?' But if the agent's job is to get through the agenda and if the manager will listen, then it could be a good way to keep meetings more on track without people having to stick their neck out with their manager."*

These quotes demonstrated that group meetings can benefit from CASSY's structure management. Whether embodiment is necessary for managing structure may depend on the nature of the discussions. I speculate that the keys lie in attention and compliance, especially when the agent suggests actions that are contrary to the individuals' current engagement. In my study, participants were generally in agreement with the agent- provided structure regardless of the embodiment—possibly because it was a laboratory experiment. The issue of whether embodiment, especially with enhanced presence and perceived social intelligence, could improve compliance with structure management in the long run merits future research. Mean- while, given that embodiment can potentially cause distraction, an un-embodied agent might be a better choice in some situations, such as in high-stakes decisions where both cognitive demand and motivation for decision success are high.

I also noticed a tension between the needs for enforcing structure and some participants' subjective resistance. While some participants praised the agent's effect on *"keeping us on track"* and *"making efficient decisions"*, a few commented on the agent being *"pushy"*, *"interruptive"*, and "*a little frustrating"*. Resolving such a tension is a critical challenge in designing meeting facilitation agents. While I found embodiment could potentially increase a sense of power, I emphasize that the non-verbal modalities may be leveraged to better exhibit social intelligence such as expressing rapport and empathy, which can potentially ease the enforcement.

Affective and social catalyst: Some participants appreciated the facilitation agent as an *"ice breaker"* in the beginning of the meeting to reduce the social awkwardness between attendees. Several groups also suggested how an agent can bring *"additional help"* for overcoming expression barriers, especially for those less confident or experienced. For example, P21 in group 11 (voice) suggested:

> *"If I was a novice, I tended to be a little more fearful about sharing my opinion like the time I suggested okay let's go back to candidate two...these are types of projections that I would expect an agent to do."*

This points to a future functional focus for group facilitation agents—to proactively mediate tensions among group members. I suggest that embodiment may help in several ways. The improvement in social perception such as rapport may bring affective benefits. The enhanced social presence may create a stronger sense of involvement to play an active social role in the group interactions. Moreover, the additional modalities such as movement and facial expression can create more triggers for affective and social responses. Future research may explore these possibilities.

Information source: Currently, CASSY provides very limited information support, such as stating how much time is left, or suggesting another criterion to consider. However, many participants expressed strong desire for group facilitation agent to provide informational support, upgrading the agent from a Level1 GDDS to Level2 [54]. Examples of desired functions include answering questions, providing reminders of the status of the decision, looking up information (e.g. definitions of terms), presenting evidence in a helpful format (e.g. a comparison table), and even providing rationale suggestions.

Moving forward, group facilitation agents providing informational or reasoning decision support may benefit from embodiment, for its exhibition of task-oriented capabilities, perceived trustworthiness and invitation for interactivity. The advantage for interactivity is especially important, because information support should not solely rely on agent-initiated conversations. Also, in a conversational form, decision support should be precise for users' information needs, so the agent may have to frequently engage in communication processes involving turn-taking, clarification, repair, etc. Embodiment has known benefits for these interactional functions [32].

4.6 Discussion

Facilitating a group meeting often requires "standing apart", taking a neutral role and avoiding the effects of organizational and personal relationships on the task. Humans can find these difficult to do. Furthermore, a small group may not wish to lose the contributions of a human member just so he/she can facilitate. These challenges in group facilitation, as well as advances in AI and conversational agents, motivated us to study how a CA can play the role of a facilitator in a group. A facilitation agent can also ensure that the same protocol to be carried out across meetings, leading to great consistency, and preventing unwanted tensions. In the following paragraphs, I discuss how my results—particularly my findings demonstrating the potential of CAs for group facilitation tasks—relate to prior work and point to design implications and future directions.

On group facilitation agent design

My study showed positive influence of embodiment on how users perceive and respond to a conversational agent in a group context. My qualitative results echoed previous work that the embodiment can improve user perceptions of: 1) the agent's social-interactional intelligence, as the non-verbal modalities make the interactions more intuitive and enjoyable; and 2) the agent's task capabilities, as signaled by a humanoid "face". I discuss some possible designs to further enhance these two aspects, specifically for group facilitation agents.

In my design of embodied agents, and in many previous studies (e.g., [16]), one key area explored is how to exhibit social intelligence through floor management (turn-taking) behaviors such as active listening, hinting at interaction intentions, and facial expressions of confusion, etc. Floor management is even more important and challenging in a multi-party setting, as the interaction dynamics become more complex. Besides the technical challenges in sensing the best timing, I suggest considering designs of agent behaviors for different floor-taking contexts. For example, in handling interruptions, the agent may first need to display cues that it is about to talk, and fol- low a protocol to handle the interruptions (e.g. immediate stop [19], lower the volume). Other contexts of floor-taking in groups include sensing disengagement, sensing confusion, and sensing conflicts. The agent can resort to learning from verbal and non-verbal behaviors of human facilitators to actively reengage, clarify and mediate the groups.

Moreover, for a group facilitation agent, it is important to con- sider how to enforce structures in a "socially appropriate" way. This requires the agent to 1) be context aware (e.g. decision status, members expertise), 2) be socially aware (e.g., group relations, power dynamics, individual preferences) and 3) ex- hibit behaviors that conform to the social norms. The last part can benefit from careful design of embodied modalities. For example, an agent that is able to convey rapport and empathy may be better received when soliciting compliance.

Enhancements in user perception of the agent's task oriented capabilities can improve both structure management and in- formational support. For example, a facilitation agent that appears "confident" may be more effective in having users follow its instructions. Therefore, having a persona that fits the task (i.e. professional), and consistently exhibiting it through visual portrayal, talking style and other non-verbal behaviors, should be fundamental considerations for the design of a CA's embodiment. It is also important to avoid creating unrealistic expectations that arise from agents' embodiments. The well-known problems of "Uncanny Valley" and "unclear affordance" of CAs suggest that a highly lifelike humanoid appearance may risk eliciting negative emotional responses [181], and creating a mismatch between user expectations and system capabilities [121]. My results also suggest that embodiment may invite proactive interactions that the system is not equipped to respond to and lead to user frustration. Future research should explore the calibration of different embodiment designs and the levels of intelligence perception they evoke.

On being present

Another unique finding from my study on the benefit of embodiment is enhanced presence. Its implications merit particular consideration for introducing CAs into collaborative roles. As some of my participants said, embodiment may create a perceptual difference between having an additional member "participating in the conversation" versus a machine "in the background". In addition to benefit from social effect such as rapport and persuasion, two additional areas of potential benefits are *attention* and *influence on group dynamics*. In some group contexts, a CA may have more reasons to compete for users' attention against other team members and the collaborative tasks at hand. For this reason, embodiment may be a preferred design. For example, besides a group facilitator, CAs can also play a teammate or a group tutor. The latter categories may have even higher requirement for continuous user attention and engagement, and can potentially benefit more from embodiment designs.

My exploratory analysis showed evidence of more balanced group dynamics with an embodied agent. The phenomenon could be caused by "social influence" created by an agent continuously "being there", and could also be attributed to more positive affect when interacting with the embodied agent. The underlying mechanism calls for future research. Understanding how the presence of an agent (embodied or not) exerts influence on group dynamics is an area worth investigation, and may have profound implications for developing CAs for collaborative tasks. Studies should look beyond synchronous collocated collaborations. For example, CAs for group chat are becoming popular applications (e.g., Slack bot). It would be interesting to study how they influence group dynamics and change how members collaborate with each other.

4.7 Limitations and Conclusion

I acknowledge several limitations of the study: 1) I could only conjecture about some of the causalities for the agent's effect on group processes. The validation awaits future research.

The "hiring decision" experimental task may impact effects I observed. For example, the lack of effects on consensus shift could have resulted from lacking real-life accountability. Future research should explore more realistic decisions. I also observed very few instances of ignoring or disobeying the agent's instructions. Whether these effects still hold in the long run is arguable. 3) As a lab study, I controlled for factors that could impact how groups react to a facilitation agent, for example, the number of participants and power dynamics. My study may also be limited by the selected demographics (enterprise employees, relative young) even though they should be the main targeted user group of the system. I acknowledge that the setup may have not observed certain group phenomena. I hope future research explore how embodiment interacts with different individual, social and cultural factors.

Conclusion

I revisited the longstanding debate on the necessity of embodiment for conversational agents in a new context–group facilitation. Consistent with previous work, I showed that the embodiment improved various dimensions of subjective perceptions of the agent, but its effect on the objective task performance was less evident. However, in the group context, I found evidence that the embodiment had positive influence on group dynamics and invited more pro-active interactions. My qualitative results suggested these phenomenon could be potentially explained by an enhanced social presence of a agent continuously "being there", more intuitive and pleasant interactions with multi-modalities, and higher task capabilities attributed to the more lifelike visual character. Although the cost of developing additional modalities does not always justify the benefit of embodiment, I suggest that embodiment is a valuable feature for CAs in collaborative contexts, especially when social influence such as rapport, trust and power is beneficial for the task, when the agent activities require continuous user attention, and when it involves collaboration and mixed-initiative interactions between human and the agent.

** This study has been conducted as a part of my internship at IBM Research. The methods and findings were published as a long paper at CHI 2018 under the following title.

"Face Value? Exploring the Effects of Embodiment for a Group Facilitation Agent" [166]

Chapter 5: Prototype II - AUTOMATED GROUP FACILITATION ROBOT

Group meetings are ubiquitous, yet many employees consider their meetings to be unproductive, dissatisfying and costly [164], due to a variety of dysfunctional behaviors, complex social dynamics, and lack of experience in conducting efficient and productive meetings. Previous studies have shown that meeting facilitators can be advantageous in helping groups reach their goals more effectively, but many groups do not have access to human facilitators due to a lack of resources or other barriers. After showing the overall feasibility of using conversational agents as group facilitator, in a Wizard of Oz study, my goal is to develop a fully automated conversational agent for group meeting facilitation. In this chapter, I describe the development of a multimodal robotic meeting facilitator that can improve the quality of small group decision-making meetings. This automated group facilitation system uses multimodal sensor inputs (user gaze, speech, prosody, and proxemics), as well as inputs from a tablet application, to intelligently enforce meeting structure, promote time management, balance group participation, and facilitate group decision-making processes. Results of a between-subject study of 20 user groups (N=40) showed that the robot facilitator is accepted by group members, is effective in enforcing meeting structure, and users found it helpful in balancing group participation. I also report design implications derived from the findings of my study.

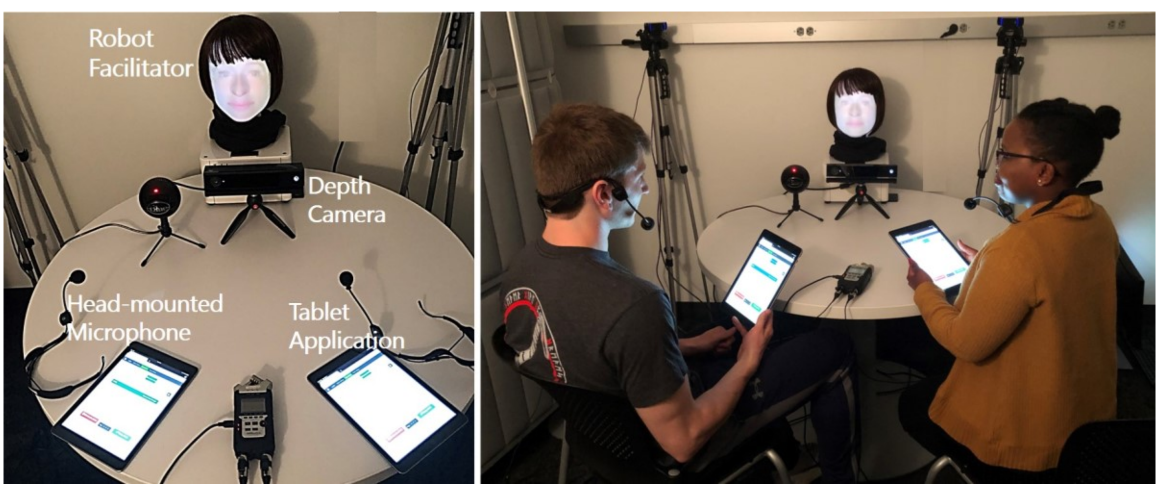


Figure 7. The automated group facilitation system and the study setting; the robot facilitates a small group decision-making session. Participants also use a tablet application for some of the decision-making tasks.

5.1 Introduction

In most organizations, meetings are one of the primary means of communication and coordination. On average, one third of employees' time is spent in meetings, involving many kinds of collaborative decision-making tasks [2, 89]. In group decision-making meetings, employees can view a problem from different perspectives, get exposure to a diverse set of insights and ideas that are not otherwise accessible, and reach consensus on decisions. However, there are many challenges in group decision-making meetings that can lead to suboptimal performance and efficiency. For example, group dynamics such as social influence, conflict, social loafing [106], groupthink [10], and dominance [30] can complicate the decision-making process and impact a group's performance. As a result, there is some debate about whether group decision making necessarily results in a more effective decision-making process and better decisions compared to individual decision making [86].

Professional human facilitators have been shown to be effective at improving group decision-making processes and outcomes. For example, Westley and Waters demonstrated that group facilitators are effective at addressing many of the common problems in meetings, such as not establishing or diverging from a meeting agenda, intragroup conflict, and domination of the discussion by one participant [197]. Viller also demonstrated that a group facilitator can greatly improve meeting quality by enforcing a structure, eliciting equal participation, and managing conflicts [189].

Despite the potential benefits of group facilitators, the high cost of hiring professional consultants or training in-house facilitators prevents many organizations from using them for the myriad meetings held every day. Automated systems may be able to perform many of the functions of a group facilitator at a fraction of the cost, and can be available whenever needed, for all meetings regardless of size or importance. I posit that a physically embodied conversational agent (ECA), with the ability of recognizing and displaying human-like interactional social cues, can provide effective meeting facilitation that requires minimal user training. I hypothesize that several capabilities of a conversational *humanoid* robot, such as the ability to interact with human collaborators in natural language and nonverbal conversational modalities, the ability to hold anon-judgmental and neutral point-of-view, and access to information resources, can make a Robot Facilitator an effective meeting facilitator. Moreover, as suggested by previous research [166] a conversational agent with human-like embodiment, functioning as the group facilitator, could improve the users' perception of the facilitator's social behavior, intelligence and authority, and serve as the locus of attention to lessen distraction. A humanoid ECA can also use speech and nonverbal behavior, such as gaze and facial display to model the behavior of human facilitators with reasonable fidelity.

Although significant research has been conducted on the design and development of ECAs that interact with single users, there is much less research on ECAs, virtual or robotic, designed for multiparty interaction. Meeting facilitation provides an ideal problem domain and testbed for multimodal multiparty interaction research that has the potential for significant societal impact.

I designed and developed an Automated Group Facilitation (AGF) system in which a robot facilitator uses multimodal sensory inputs to moderate a small group decision-making session. In my system, meeting facilitation functions, such as agenda management, time management, and participation management, are layered on top of more foundational conversational processes, such as turn-taking and grounding, which are necessarily multimodal. My robotic facilitator uses cues from user gaze, speech, prosody, and proxemics, in addition to task status from a tablet application, to coordinate a decision-making meeting.

In this chapter, I describe a prototype multimodal robot meeting facilitator, that performs the following basic functions: agenda management, to ensure all essential topics are discussed; time management, to ensure the meeting is conducted in an efficient manner; and participation management, to ensure the opinions of all participants are heard. I evaluate my prototype in the context of a hiring meeting, addressing some of the known challenges in these meetings, including lack of structure and imbalanced participation (dominance). In addition to measuring task outcomes, I am particularly interested in assessing whether participants accept an ECA in the role of a facilitator, and how an ECA impacts the social dynamics of meetings.

5.2 Related Work

5.2.1 Technology-Driven Group Facilitation

Viller examined the role and responsibilities of a group facilitator in computer supported cooperative work (CSCW) and discussed the extent to which this role should be automated in the group setting. He described how a facilitator's role changes over a meeting lifecycle, from a more central role in the beginning, stepping back and only providing high-level interventions in the middle stage, and again taking on a more directive role at the end, integrating the ideas and wrapping up the meeting. Viller identified five general meeting problems that can be addressed by a group facilitator, including: lack of structure, domination, conflict, and lack of engagement [189].

A number of Group Decision Support Systems (GDSSs) have also been designed to support or even replace some of a human facilitator's roles. GDSSs in general can support face-to-face and remote group meetings and different types of GDSSs are designed to remove communication barriers, enforce structured decision making techniques, and use formalized rules to systematically direct the timing or content of discussion [53].

To tackle a lack of leadership and promote structured collaboration in distributed groups, Farnham et al. customized a chat system to enforce the structure of group communication and the group decision-making process. They conducted an evaluation user study of the system in a hiring decision making scenario, and demonstrated that participants who used the system were more likely to reach consensus, made higher quality decisions, and showed a greater recall of discussions [62] .

Prior work in the CSCW community has also investigated how technology can enhance group collaboration and improve unequal participation in meetings. For example, Meeting Mediator is a real-time portable system that provides visualized feedback of members' verbal participation on a mobile display to enhance overall group collaboration [111]. Furthermore, much research has been done to design smart meeting rooms both to provide real-time assistance and to capture human behavior in groups [36, 190]. Smart meeting rooms are usually equipped with different sensors, multiple cameras and microphones to capture details of a meeting and detect contextual information. In a recent work by Bhattacharya et al., a model is proposed to automatically process multimodal sensors outputs, and detect human verbal and non-verbal behaviors and understand group dynamics [14].

The work to date in this area indicates the great potential of employing intelligent systems for group facilitation. However to the best of my knowledge none of the work on GDSSs has leveraged multimodal ECAs as virtual meeting facilitators.

5.2.2 Robots and Conversational Agents in Groups

Advances in robotics technology have enabled a growing number of robots designed to work in human teams [78], with examples spanning from invasive telesurgeries [60] to space exploration missions [58]. Several researchers have studied how the presence of robots in teams can affect human group dynamics and team performance. Jung and his colleagues have explored the use of robots to moderate team conflicts through a Wizard-of-Oz study in which a robot attempts to repair interpersonal violations in 3-member teams [101]. Jung also introduced the concept of "affective grounding", as one of the essential prerequisites for coordinating human-robot interaction [99]. Matsuyama conducted a series of studies investigating a robot-based facilitation framework designed to balance participant engagement in a group conversation. Results of an evaluation study indicated that the procedural approach to manage the floor was more accepted and improved feeling of groupness [125]. Robots have been also used as moderator and team member in the game context to study the effects of different roles that a robot can play in game teams [7, 25]. Other studies investigated effects of a conversational agent's embodiment in the group setting and found physical embodiment of robots can improve groups engagement, attentiveness, and social perception of the agent [3, 166]. Tennent et al. studied effects of a non-humanoid robot promoting participant engagement in a group. They evaluated a peripheral robotic object (microphone) that shows non-verbal implicit actions to follow and encourage participation in a group conversation [179]. Their intervention resulted in increased participant engagement.

There is also a growing body of literature on ECAs designed for multiparty interaction [6]. For example, Bohus and Horvitz conducted a series of studies and proposed a model to represent a CA's turn-taking behaviors by gaze, gestures and speech [18]. Such a model could smooth the ECA's speech flow in a multiparty conversation. Skantze et al., also explored how a robot can manage turn-taking and attention when multiple players are working with a shared display [171]. In other studies, non-verbal cues such as gaze and respiration were used to predict the next speaker and facilitate turn-taking management in a multiparty interaction [91].

5.3 Automated Group Facilitation (AGF) Robot; System and Architecture

The Automated Group Facilitation (AGF) system is designed to support collaborative activities in group meetings by providing interventions to mitigate some of the known challenges such as diverging from agenda and domination. The AGF system is multiparty, interacting with all meeting participants, and is multimodal, using diverse inputs and outputs to support natural conversational interaction with group members. To further facilitate natural interaction, the system is personified with a *humanoid* robot that serves as a group facilitator (Figure 1) and is able to conduct a quasi-natural conversation.

Several studies have demonstrated that physical robots are more engaging than virtual (screen-based) agents [3, 166], and the robot's physicality improves user's perception of its presence in the team [36]. A physical robot can also use its gaze to more accurately display attentiveness and feedback to speakers and to indicate a desired addressee [2]. Thus, I chose to use a physical robot as the group facilitator.

More specifically, the AGF system uses three channels of sensory inputs to enable the robot to see group members (via 3D depth camera), hear their voice (microphone), and understand the decision-making task status (tablet application). Multimodal outputs of the system include: the synthetic speech of the robot, conversational nonverbal behavior (lip synchronization, communicative facial displays, gaze orientation), and actions such as updates to the task application on tablets.

The main goals of the system are to improve overall meeting structure and quality and increase engagement in and satisfaction with the collaborative activity by providing just-in-time instructions. To achieve this, the system is designed to provide different types of support during a group decision-making session: social facilitation, and decision-making and meeting facilitation. Besides the facilitation services, the system needs to be equipped with technical capabilities to support a multiparty interaction. In the following paragraphs, I describe each of these facilitation types in more details.

Social Facilitation

One of the most important tasks of a group facilitator is to catalyze the social interaction among members of a workgroup [165]. Social facilitation tasks of a group facilitator include: greeting everyone and introducing herself and the members of the group; "setting the stage" by orienting group members to the task and process; providing "ice-breakers" to acquaint group members with each other and begin building trust and rapport; discussing desired meeting outcomes; and reviewing the agenda. For group members to accept the robot in this social role, it needs to be able to engage them in natural, multimodal, face-to-face conversation. In order to build rapport and trust as a social facilitator, it needs to display understanding and empathy using conversational backchannels [13] and "active listening" behavior [118].

Decision-Making and Meeting Facilitation

In order to alleviate some of the common challenges of group decision-making meetings, such as inefficient and unstructured discussion and imbalanced participation, the AGF system should be designed to provide five types of facilitation intervention:

- **Enforce meeting structure:** To avoid off-topic discussions and deviation from the meeting agenda, the system should be able to enforce the meeting structure by offering step-by-step instructions and monitoring the decision-making process. The system needs to know the context and the meeting stage to offer timely feedback, and guide the meeting flow by asking targeted questions.
- **Encourage and Balance Participation:** Dealing with people who tend to dominate the conversation in a meeting is an important function of a group facilitator [134]. To ensure balanced participation, and to create a safe space for all participants, the system needs to detect when some participants are significantly less vocal during a meeting and engages them in the conversation.
- **Content-based Feedback:** A group facilitator should also provide feedback related to the task being discussed by the group. Hence the AGF system needs to recognize some speech content to provide interventions based on the topic of participants' discussion. For example, the system should track aspects of the task that need to be discussed and remind the group if they forget to discuss them.
- **Manage the Time:** Group facilitators keep meetings on track by watching the time and agenda. The AGF system should also track the time spent on each task in the meeting to both remind team members of the time, if needed, and offer comments based on the time and the stage of the meeting.
- **Maintain Authority:** A facilitator should be perceived as an authorized and intelligent member to enforce participants attendance and adherence to her instructions. As suggested by Justine Cassell, the computer agent's embodiment helps users to better locate its intelligence and power and improves the perceived trust [34]. Using a conversational robot with a human-like face and nonverbal behaviors such as gaze and facial expression as the

group facilitator, allows participants to treat it as an intelligent and authorized social entity with a persona.

Multiparty Conversational Competence

Another important feature of the group decision-making setting is the involvement of multiple humans in the interaction, so the Robot Facilitator should be able to support multi-party conversations. In order to engage in any natural interaction with group members, the robot needs to have certain minimal conversational competencies. To conduct multiparty interaction with all group members, the robot is designed to receive inputs from multiple sources in order to detect the speaker and addressee of any utterance. It also needs to understand at least some user speech content to provide the right action or response via speech recognition. The robot also has minimal conversational turn-taking management capabilities, tracking which participant is speaking and indicating attentive behavior by gazing at the speaker, as well as indicating her intention for a specific user to speak by gazing toward them.

5.3.1 System Architecture

To satisfy the requirements mentioned above, I use a 3-layer system architecture to model group interactions, recognize participant behavior, and provide proper intervention by the robot. This framework adapts the Smart Meeting Rooms' architecture [190] and integrates it for use with a conversational robot. In this architecture, audio data and tablet application inputs are used to track the meeting stage, calculate participation time, and provide proper interventions based on the meeting stage, time, and content. Visual data from the camera is processed to make the robot aware of the environment and generate adequate non-verbal and verbal listening behavior. The components are controlled in different software modules, and I use an Event-Driven Architecture to handle messaging between modules and the dialogue manager. Events from the camera and microphones are transferred in an internal framework (described in section 4) and events from the application are stored in a local databased and passed to the dialog manager through the network. Figure 3 shows the architecture.

5.3.2 System Components

Here I briefly describe each of the main modules of the Automated Group Facilitation architecture to highlight how they were used to handle an interactive multiparty dialogue and satisfy the requirements mentioned above. Figure 1 shows the deployment of the AGF system components in the experimental setting.

Humanoid Robot

The facilitator robot is a humanoid head developed by Furhat robotics (Figure 1). It has an animated face, back-projected on a translucent mask, mounted on a two degree-of-freedom mechanical neck that allows the head to turn and nod [3]. The robot's speech is generated using the Cereproc speech synthesizer. The robot can direct its attention using eye gaze and head pose, show emotions or listening behaviors by moving her eyebrows, lips and nodding her head.

3D-depth Camera

Participants' location, and head positions are captured using a Microsoft Kinect camera. To generate appropriate non-verbal responses such as listening behavior or gazing at the speaker, the AGF system used this input to identify participants' gaze direction using their head orientation.

Head-mounted Microphones

The AGF system obtains audio data from microphones worn by each participant. The audio input is passed to an automated speech recognition (ASR) module used by the dialogue manager.

Tablet Application

I designed and developed a custom tablet application to support the group decision-making tasks. The application can be used for individual and group activities during the group-decision making activity. It provides some of the basic features of group decision support systems such as voting and sharing the decision status. The application facilitates the communication and sharing of opinions by providing a discussion interface that reflects the group decision at each stage of the decision-making process. This application provides another input channel for the AGF system. The AGF system used this input in addition to the speech input to better detect the current topic, the meeting stage and the decision status.

5.4 Task Domain: Hiring Meeting

Although my system architecture and methods can be used for most kinds of group decision-making meetings, I grounded my development in a single task domain: a hiring meeting. The hiring decision meeting is a very common scenario for group decisions in the workplace, and previous work used similar tasks to study group decision processes [62]. This experimental task and procedure are adopted from previous work [166], in which the authors explored the effects of embodiment of a virtual facilitator in a group setting. For this task, a group of two participants are asked to review and discuss a set of six resumes for a sales manager position, and select the one best candidate for an interview. I created fictional resumes based on existing online resumes of applicants for the same position. Resumes were designed so that the candidates each had unique strengths and weaknesses and did not have an obvious rank ordering among them for the position to be filled. Participants are given 30 minutes to review, discuss and select the best candidate. An application installed on tablet computers provided to each group member (Figure 7) to perform some of the group decision-making tasks and review the resumes. A humanoid robot served as the facilitator of the group decision-making session.

I designed a standard agenda [129] to structure my experimental hiring meeting (Figure 8). The steps in this agenda are:

1. **Greeting and Introduction:** The facilitator initiates the conversation by introducing herself, inviting the participants to introduce themselves, and greeting them with their names (she greets back if participants ask her). The facilitator then leads several "ice-breaker" exercises to prepare the group for the task. For example, she asks each participant to tell her and the other participant about their favorite part of their work. She can also ask participants whether they knew each other before the hiring meeting. She then asks participants whether they are comfortable with the setting for the meeting. This conversation acts as a social catalyzer to build rapport and trust among group members.

2. **Agenda Description and Resume Review:** The facilitator describes the goal of the meeting and the agenda, then asks the group to review hiring candidate resumes individually for a few minutes and then provides a preliminary ranking using the tablet application. She also suggests what criteria they should consider in their initial review, and describes how the tablet application interface should be used to rank the resumes. She lets participants know that she will keep track of the time, reminds participants as the time runs out, and can negotiate time extensions if participants require more.
3. **Criteria Discussion:** When participants are ready to move to the next agenda item, the facilitator suggests that the group discuss the hiring criteria. She prompts participants to consider the job description, if it has not already been discussed, then she waits for the group

to discuss the main criteria on the resumes (e.g., education, work experiences and skills). Once these have all been discussed, the facilitator introduces the next agenda item. If participants indicate they are ready to move on but have forgotten to discuss one or more criteria, the facilitator prompts them (e.g. *"What do you think about their education?"*). During the discussion the facilitator provides acknowledgements and active listening behaviors. She also monitors vocal participation and prompts less active participants by name during breaks in the conversation.

4. **Main Discussion and Elimination of Unfavored Candidates:** The facilitator invites the group to start discussing the hiring candidates, by going through all of the resumes and eliminating the unqualified ones. When participants converge on a decision (by indicating in the tablet application they are keeping or eliminating a particular candidate), the facilitator confirms that decision. If participants skip to another candidate too soon without a decision, the facilitator asks them whether they want to make a decision first. If they cannot decide about a candidate after a few minutes, Sarah suggests that the group discuss the next candidate. After all candidates are discussed, if too few have been are eliminated, the facilitator suggests that the group discuss and eliminate more before proceeding. Once the hiring candidates have been narrowed down sufficiently, the facilitator summarizes the remaining candidates, and moves on to the next agenda item. As in the previous task step, the facilitator also monitors the time and vocal participation.
5. **Selecting the Best Candidate:** When the number of hiring candidates has been sufficiently narrowed, the facilitator asks participants to discuss the pros and cons of the remaining candidates to select the best one. If participants have difficulty (pause for too long), the facilitator provides suggestions to facilitate the decision-making (e.g. to review the criteria mentioned in the job description, or suggest the group reflect on their initial voting, or compare them with each other). The facilitator again monitors time and vocal participation.
6. **Decision on the Final Candidate:** When the overall meeting time is up, or the group reaches a final consensus decision, the facilitator summarizes the session and confirms the final decision with each participant.

7. **Wrap up the decision-making task:** Finally, the facilitator thanks the group for participating and asks them to use the tablet application to rank the candidates individually based on their discussion.

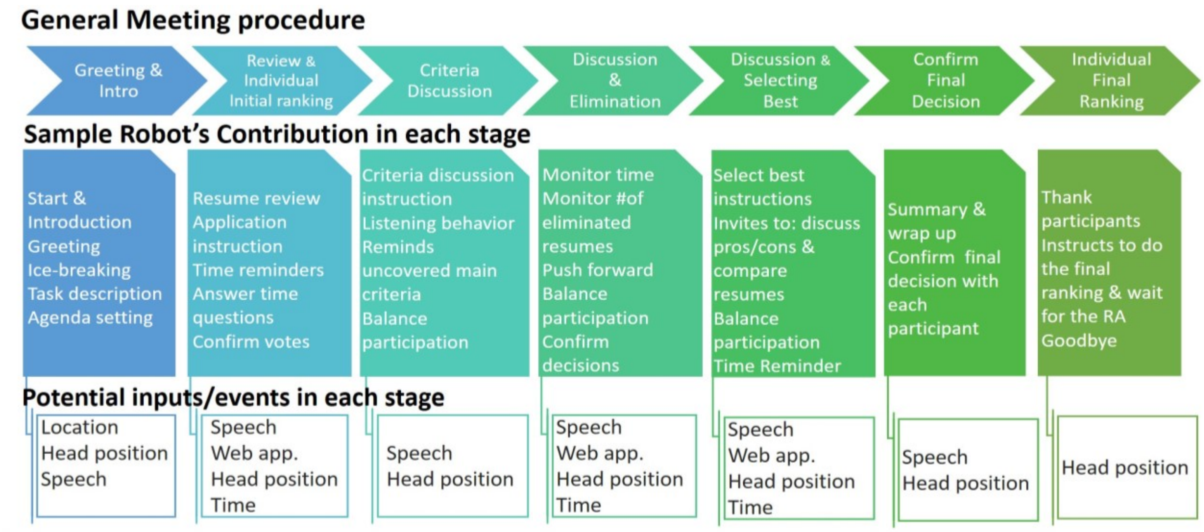


Figure 8. Group decision-making procedure. Examples of robot's contribution and multimodal inputs in each stage are shown in the bottom.

5.5 Dialogue Management

The dialogue manager of the AGF system is designed to accomplish five main goals:

- Handle multiparty interaction
- Enforce structure for the group decision-making meeting
- Ensure content coverage by providing content-based recommendations
- Balance group participation by nudging the less active participants
- Track and manage the time by conveying rule-based guidance

I used IrisTK, an open source dialogue system framework [170] to manage the conversation flow in the multiparty interaction with the Furhat robot, using a state machine-based dialogue manager. Using IrisTK I designed an event-driven rule-based dialogue manager that uses information from the input modules and creates appropriate responses. In section 5.1 and 5.2, I describe more details on the input signals, the output actions the system provides to achieve the goals mentioned above, as well as the protocol through which the system uses inputs to generate appropriate outputs.

5.5.1 Input Modules

Four main input modules are used to capture the meeting.

Visual input from depth camera: A Kinect depth camera was used to capture visual and depth information. Inputs from the Kinect enable the system to sense user presence, identify user location, and estimate their gaze direction using head orientation as a proxy. This module provides the number of users in the scene, sends an event when a user enters or exits the scene, and detects whether the user is attending to (gazing at) the robot or not.

Voice input from microphone: User speech is captured by head-mounted microphones and recognized and converted to text using Google's cloud-based automatic speech recognizer (ASR). The speech signal is also processed to provide three additional types of information: 1) silence vs. speech intervals for each user; 2) who the current speaker is; and 3) user vocal participation from the accumulated speech time of each participant in each stage of the meeting. To understand the output of the ASR module, and to handle the semantic interpretation of the speech input, I used the W3C Speech Recognition Grammar Specification (SRGS)[6] and Semantic Interpretation for Speech Recognition (SISR)[7] tools integrated into the IrisTK framework. I adopted an XML form to represent context-free grammar rules. The recognized semantic intent of the speech input is used to drive robot behavior in each dialogue state throughout the interaction. For example, using SRGS and SISR allows the system to detect more general speech input such as answers to yes-no questions, repeat requests, queries including the robot's name, questions and thanking utterances.

---

[6] https://www.w3.org/TR/speech-grammar/
[7] https://www.w3.org/TR/semantic-interpretation/

Decision status from the tablet application: Participants use the tablet application to review and rank candidates individually, and to share ideas during the group discussion. The application has three views for the three stages of the meeting: 1) Individual resume review and initial ranking (Fig.9.a) 2) Group discussion, elimination and selecting the best candidate (Fig.9.b and c), and 3)

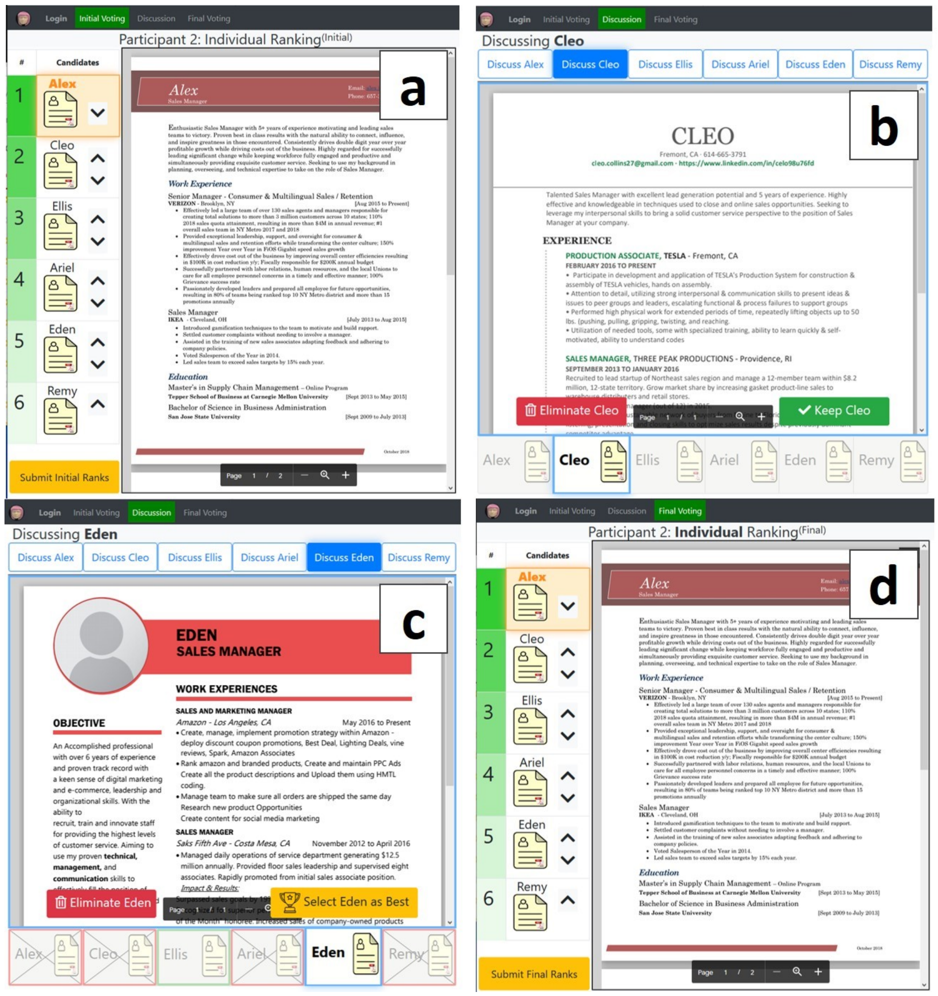


Figure 9. The web application; a) Initial individual ranking interface; b) Group discussion interface ;c) Group discussion interface after eliminating four candidates; d) Final individual ranking interface

Individual final ranking (Fig.9.d). During the group discussion phase, most of the participant actions are shared with other participants to facilitate communication and sharing of ideas. Figure 4 depicts various views of the tablet application. Each participant is given a touch tablet to use the application. The application logs some contextual data (e.g., the meeting stage, the active candidate, and the group decision status on each candidate) and sends signals to the dialogue manager on occurrence of each event. The dialogue manager obtains data such as the user names, each candidate's status, and the number of discussed and eliminated candidates from the application. Other input signals are received from the application at the four following moments during the session: 1) when initial ranking is submitted during the individual review phase, 2) when the active candidate is updated during the group discussion phase, 3) when the decision status on each candidate is updated (e.g. kept or eliminated), 4) when the best candidate is selected.

Time: In the dialogue manager, timer events are used to initiate timing signals at several occasions during the interaction. Examples include: during the resume review phase to send time reminders, during the discussion phase to track the time spent on each stage and on each resume.

5.5.2 Output Actions

In this section, I describe the outputs generated by the dialogue manager based on inputs presented in the last section. The three different types of outputs are: non-verbal, verbal, and application actions.

Non-verbal behaviors

The robot facilitator performs a range of human-like non-verbal behavior to show active listening behavior and to manage the turn-taking in multiparty interaction. When the group members enter and when they are conversing, the robot gazes towards the active speaker, smiles, nods her head, and uses confirmation filler words (aha, I see, that makes sense, etc.). Across all stages of the meeting, the robot handles multiparty interaction by attending to (i.e., gazing at) each participant when she asks them a question, or when they are speaking. The robot looks at the group (i.e., shifts gaze among participants) when she is making a general statement.

Verbal behaviors

Using the processed information from the input modules the dialogue manager decides what the robot facilitator should say based on the meeting stage, user speech input, time, or inputs from the tablet application. The verbal outputs are categorized as follows:

1. Robot-Initiated Utterances: These are the instructions, comments, and questions that the robot utters based on the dialogue state, the meeting stage, and/or the decision status. Examples include: asking each participant to introduce themselves at the start of the meeting; providing the task description and goal to be accomplished at the beginning of each stage, or suggesting the elimination of candidates during the discussion phase of the hiring meeting.

2. User Speech-Initiated Utterances: Participants are told that they can interact with the robot at any time in a way similar to how they interact with humans. The robot is designed to respond to participants when they call her name, greet her, ask a question about the time, thanking the robot, etc., and any point during the meeting. The system uses the interpretation of the greeting responses to provide a proper greet-back utterance or tracks which criteria have been covered by understanding the discussed criteria. Using SRGS and SISR, the system can interpret and respond to time-related queries such as time extension requests or questions about how much time is remained.
3. Tablet Application-Initiated Utterances: These commands are triggered as users interact with the application on their tablets. For example, when a participant submits a candidate ranking on the application, the robot thanks them and check the status with the other participant.
4. Time-based Utterances: Using the internal time of the system, the robot provides time reminders at certain points during the session. For example, when the participants are reviewing resumes, the robot reminds them of the time at key moments and guides them to focus on the main points for initial ranking.
5. In all dialogue states, if the robot does not hear users speech input (the voice is low), or if she does not understand the input (e.g. ASR output does not match expected semantic intents), in response to a query, she informs participants that she did not understand what they said and repeats her question. Also, in cases that the system receives multiple speech input at the same time, the robot handles the responses separately.

Actions

The dialogue manager may also output an action instead of verbal or non-verbal behavior. For example, the dialogue manager may extend the time allowed for a given decision-making task (upon user request), move forward to the next task, update the tablet application interface, or calculate participants' contribution time to nudge less active participants.

5.5.3 Meeting Facilitation Services

The AGF system achieves the Decision-Making and Meeting Facilitation functions described in Section 4 as follows.

Enforce meeting structure: The system tracks the meeting stage and enforces the meeting structure by providing certain services such as offering step-by-step instructions to move forward to next stage, asking targeted questions, acknowledging actions and asking confirmation questions, monitoring the decision-making process and suggest corrective actions (e.g. try to eliminate more options), providing feedback, decision summaries and wrapping up the sessions.

Ensure content coverage: At some stages during the meeting, the system listens to the discussion and provides comments and instructions based on keywords detected in the content. For example, the system monitors the conversation when participants are discussing hiring criteria, and reminds them to further discuss any of the main criteria that have not been adequately covered. The system analyzes the ASR output and uses an off-the-shelf keyword detection (SRGS) module to provide content-based interventions.

Encourage and Balance Participation: At the end of each stage of the decision-making process (e.g. criteria discussion or each resume discussion), the system references the tracked vocal participation for each user as their contribution to the discussion. If an imbalance is discovered (e.g. >10 seconds difference in speech contribution), the robot nudges the less active participant by asking what s/he thinks, or if s/he wants to add anything else.

Manage the Time: The robot provides some comments and feedback based on the time and the stage of the meeting, as ascertained from the tablet application input. The system also tracks the time spent on each stage, and is able to extend the time upon request.

5.6 System Evaluation

I conducted a randomized between-subjects experiment to evaluate the effectiveness of the automated group facilitation system used by teams of two participants in the hiring decision-making task described in Section 4. I compared the full system (the robot, tablet application, and supporting multimodal system) to a control condition in which participants performed the same hiring decision-making task without the facilitation robot (they still had access to the same tablet application for individual tasks and group decision-making). In both conditions, each participant was given a job description printed on paper, and a tablet computer (iPad) with the

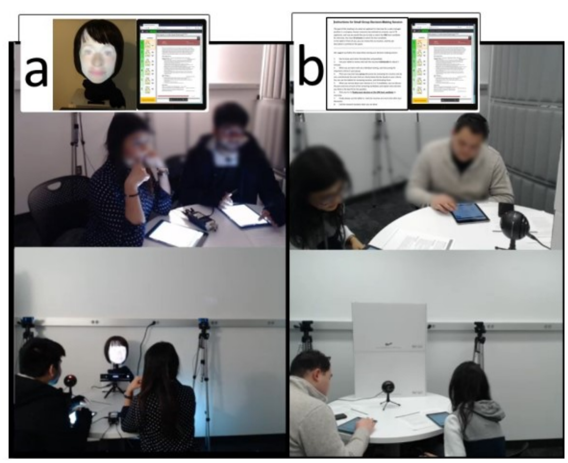


Figure 10. Experimental setup for the evaluation study. a) robot condition; b) no-robot condition

hiring application running. Participants used the application for their individual review and ranking of resumes, as well as to support group discussions. In the ROBOT condition, the robot facilitator provided instruction for each stage of the meeting (see Section 4)]. In the CONTROL condition participants were provided with written instructions for the meeting, including the same agenda that the robot uses in the ROBOT condition, as well as time management instructions. Figure 10 shows the setup of the experiment. All sessions were video recorded using four synchronized cameras for subsequent analysis.

The tablet application was used in both conditions to view the resumes of hiring candidates, and to record the results of individual candidate ranking and display group decisions (Figure 4). In the ROBOT condition the interface was automatically updated based on the meeting stage and intervention actions taken by the AGF system, while in the CONTROL condition navigation instructions were shown as alert messages, and participants could use a navigation bar to move to different views.

Participants: Participants were recruited by posting advertisements around our institution and on online job posting sites. Participants were required to be 21 or older, be able to read and speak English, and have some professional work experience involving participation in professional meetings. Participants were randomly matched to form two-member groups. To minimize the effect of gender on group performance, I ensured that the ratio of mixed-gender/same-gender groups were equal in the two conditions. The study was approved by our institution's IRB, and participants were compensated for their time.

Measures: I measured both subjective responses from questionnaires and objective measures reflecting the individual and group performance in the decision-making process. Users' perception of rapport with the facilitation robot was assessed by asking a set of questions suggested by [146]. Attitudes towards the robot were assessed in the ROBOT condition using the bond subscale of Working Alliance Inventory [88], and single item questions about the robot (e.g. the robot's knowledgeability, authority, and friendliness), and perceptions of the its effectiveness in managing participation and disagreements in the group (Table 1). Participants rated their feeling towards their team and the decision-making process with a 5-item 7-point Likert scale (items: '*The members of this team get along well together*', '*I am pleased with the way my partner and I participated in the decision-making task*', '*I am very satisfied with working in this team*', '*I felt my* *partner and I understood each other*', '*I found we reached a decision efficiently*'). Perceptions of task conflict were also measured using a three-item scale introduced and validated by Jehn et al., [75] (items: *'How frequently did people on your team disagree regarding the work being done?', 'How frequently were there conflicts about ideas on your team?', 'How frequently were there differences of opinion on your team?'*). Evaluation of the overall system was assessed via a 9-item Likert scale questionnaire that included five questions to measure general usability (sample items include: *"I am satisfied with the interaction with the facilitation system", "I found it easy to interact with the facilitation system."*) as well as four questions assessing decision-making facilitation usability of the system (sample items include: *"The facilitation system was able to manage the discussion and interaction during the meeting."*). Lastly, I asked participants to score their perceptions of group performance out of 100.

Confidence in the hiring decision was assessed via self-report within the tablet application (1=not at all confident, 5=very confident), measured immediately after participants performed their individual rank-ordering, and after they reached a final hiring decision as a group. Task outcome was also assessed by analyzing initial and final candidate rankings. Metrics of meeting structure and time management were derived from analysis of session video recordings, coding whether groups performed each of the task stages, and how much time they spent on each stage (Table 2).

Procedure: Prior to the study session, a research assistant described the decision-making task, obtained informed consent, and administered sociodemographic questionnaires. Participants were given with the tablet application. The research assistant then showed them the tablet application, explained the interface, and how to use it for different tasks. Participants were told how much time they had for the task and were guided to the room with the facilitation robot to start the experiment.

In the experiment room, the research assistant first introduced the robot and explained how she will see and listen to the participants in the ROBOT condition. Participants were instructed to wear head-mounted microphones and seated, and the research assistant left the room after starting the program. The research assistant monitored the session from another room and came back to end the session when either 30-minutes had elapsed, or the group reached a consensus hiring decision. Then participants were then asked to fill out post-session questionnaires, and a semi-structured interview was conducted with both team members. At the end, the participants were given $20 compensation as well as a debriefing document disclosing the study goal and the other study condition. In total the session and interview took 90 minutes.

5.6.1 Quantitative Results

A total of 40 individuals (20 teams) completed the study. Participants were mostly (60%) male, aged 24.2 (SD=1.8). Most (95%) participants did not know each other before the experiment, and most (85%) were university students. Among the 20 groups, 11 (N=22) were randomly assigned to the ROBOT condition, and 9 (N=18) assigned to the CONTROL condition.

I analyzed participants subjective perception of the facilitation robot, as well as their perception of the group-decision making experience as reflected in self-reports. Participants in the ROBOT condition rated the robot significantly higher than neutral on *rapport*, t(21)=10.08, p<.001, and *Working Alliance*, t(21)=8.08,p< .001. Participants also rated the robot's knowledge, authority, friendless and her abilities in meeting facilitation significantly higher than neutral (Table 5).

Table 5. Self-Report ratings of the robot facilitator (Sarah) (N = 22), single sample Wilcoxon signed ranked test demonstrates all ratings were significantly greater that neutral

Items 7-point Likert scales (1= strongly disagree, 4=neutral, 7= strongly agree)	Robot Mean (SD)	p-value
I found the facilitation robot **knowledgeable**.	5.09 (1.7)	.014
I felt the facilitation robot was **powerful and confident**.	5.45 (1.4)	.001
I felt the facilitation robot was **friendly and warm** toward me.	6.36 (0.7)	.000
The facilitation system assured that my partner and I have **equal chance** to express our ideas.	5.86 (1.4)	.000
The facilitation system allowed **sufficient discussion**.	5.27 (1.6)	.003
The facilitation system **encouraged** participation	6.27 (1.03)	.000
The facilitation robot could effectively **manage the disagreements** between my partner and I.	4.95 (1.4)	.016
The facilitation robot **attempted to manage the conflicts** raised in the meeting.	4.81 (1.4)	.021

An independent sample t-test on responses to the team and decision-making satisfaction questionnaire (α=.88) indicated that participants in the ROBOT condition (Mean (SD)=6.4(.56)) were relatively more satisfied with their team and decision-making activity, compared to participants in the CONTROL condition (Mean(SD)=5.8(1.2)), t(38)=1.8, p=.06.

I also found groups in the ROBOT condition (Mean (SD) =1.8 (1)) reported less task conflicts than those in the CONTROL (Mean (SD) =2.5 (1.2)), and independent sample t-test showed marginal significance ($t(38)=-1.69$, $p=.09$.

I did not find any significant difference between the ROBOT condition (Mean (SD)= 5.68(1.05)) and the CONTROL condition (Mean (SD)= 5.4(1.48)) evaluating the system usability regarding decision-making facilitation as reflected in a 4-item composite measure (α =.81). However, when I examined the general usability of the overall facilitation system (via a 5-item scale (α =.95)), I found participants in the CONTROL condition (Mean (SD)=6.08 (1.1)) rated the tablet application significantly higher than individuals in the ROBOT condition (Mean (SD)=4.99 (1.6)) , $t(38)=-2.42$, $p=.02$. There were no significant differences between the two conditions on perceived performance and dominance.

There was no significant difference in the time (minutes) taken to reach a decision (ROBOT: Mean (SD) = 25.5(5.8), CONTROL: Mean(SD) = 24.7(5.2)), nor differences in the final selected candidate between the two conditions. However, I did see a significantly greater increase in decision confidence in the ROBOT condition (Mean(SD) =1.32(.9)) compared to the CONTROL condition (Mean (SD)=.56(.5)), $t(38)=2.94$, $p=.03$, with change measured immediately after participants performed their individual rank-ordering, and again after they reached a final hiring decision as a group.

Analysis of session videos yielded several significant differences in group behavior between conditions for each stage (Table 6).

Greeting: among the nine groups in the CONTROL condition, three groups did not have any forms of greeting and three groups had a very short (~10 seconds) introduction. All participants in the ROBOT condition greeted each other and the robot.

Criteria discussion: None of the groups in the CONTROL condition discussed hiring criteria before they started discussing candidates, while all groups in the ROBOT condition did this.

Final decision: In the ROBOT condition, after participants discuss the pros and cons of the final candidates, they select the best candidate, and then the robot confirms the consensus decision with each participant. In the CONTROL condition, the session usually was ended by one of the participants selecting the best candidate; only three groups (33%) reviewed and confirmed the best candidate together.

Participants in the CONTROL condition spent significantly more time reviewing resumes on their own before beginning discussion, t(17)=-4.27, p=.001. I also found a common pattern in the CONTROL condition in which one participant finished their review and waited in silence until the second participant finished his/her ranking. This caused the individual review stage to be significantly longer in the CONTROL condition (Mean (SD)=109.3(83.4)) compared to the ROBOT condition (Mean (SD)=24.8(27.8)) in which the robot confirms when one participant submits their individual ranking and asks the other participant whether they are finished or not, t(17)=-3.02, p=.008. During the Discussion stage, all groups in the ROBOT condition reviewed resumes in order (as shown in the application), compared to only 33% of those in the CONTROL group.

Table 6. The percentage of groups that performed each suggested meeting stage, and the amount of time spent on each stage, in each condition.

Meeting Stage	Percentage of groups who performed this stage			Time spent by groups on each stage (in seconds)		
	ROBOT (N=10)	CONTROL (N=9)	chi-square p-value	ROBOT (N=10) Mean (SD)	CONTROL (N=9) Mean (SD)	t-test p-value
Greeting	100%	67%	.03	152 (21)	21 (43)	.000
Individual review and ranking (Initial)*	100%	100%	1	307(56)	639(239)	.001
Criteria Discussion	100%	0	.000	172(38)	0	.000
Group Discussion	100%	100%	1	833(226)	598(302)	.06
Select Best Discussion	100%	89%	.82	91(85)	87(109)	.94
Finalize and Confirmation	90%	11%	.001	25(11)	1(2)	.000

* This is total time that the group spent on resume review and initial ranking phase. If participants finished their review in different times we considered the longer time here, as the group could only move forward when both participants finished their ranking.

5.6.2 Qualitative Results

From my qualitative analysis I derived themes relating to the advantages and disadvantages a robot facilitator can bring to a small group decision-making setting.

Overall most of my participants found the robot facilitator *"helpful"*, *"friendly"*, *"interactive"* and *"well designed"*. All groups said that they would use and recommend using this system for group decision-making meetings, but some groups found it especially more helpful in large and diverse groups when it is more likely to have disagreements, domination, and off-topic conversation. Several groups also mentioned that to them the robot "*is not only a robot*" and they felt "*there is a third person sitting*" with them, who unlike humans is very "*unbiased*".

The Robot's Role in Balancing the Participation

Out of 11 groups in the ROBOT condition, 10 groups proactively talked about how the robot "*ensured equal participation from both [participants] and gave [them] equal opportunity to speak up*". Many people appreciated this feature by comparing this experience with their past group meeting experiences. For example, group 8 discussed it like this: "P1: *Group discussions mostly are somewhat chaotic. [but] here we were given a fairly good chance to speak about each and every resume and to state what we like and what we do not like .... The opportunity given here was fairer than what generally happens in group discussion.* P2: *and the usual respect that we need to use space for each person is ignored, but Sarah*[8] *maintains that well.*". Some groups also found this feature helpful for resolving conflicts in the teams. For example, P1 in group 13 said: "*She gave Equal opportunities to both of us to speak, that kind of resolved conflict itself because everyone feels their voice is being heard."*

The Robot's Role in Organizing the Meeting and Time Management

My participants found the robot helpful as a meeting moderator who follows a protocol step by step and guides the team. Participants mentioned how they found the robot's guidance helpful in remaining on track: *"[G15-P1] if we could not eliminate more than two people she was encouraging that we go through it again and come to a conclusion. So this part was really good. [G15-P2]: she was trying us to engage in conversation and have discussion, and she was guiding the whole process."*. One group compared their past experiences in group meetings and mentioned how the robot made the experience different: *"Sarah would be more efficient when we want to ensure that we are not digressing from the agenda. Because from our past experiences we easily identify people we who usually have the habit of not speaking on the topic. So Sarah can be good … because she keeps everyone on track. I would defiantly recommend Sarah to be used for that."*. Moreover, all groups reported on the usefulness of the robot in keeping track of the time, however some people desired less frequent reminders even when they confirmed they would have spent more time on the task without the time reminders.

---

[8] Sarah is the name of the robot.

Several groups mentioned the slow interactional response of the robot and delayed reaction as the worst thing about the system. "*the only problem with this one was it was taking a lot of time because of the delay. Otherwise the experience was really good." [G19-P1]*. Two groups indicated they would have preferred the robot to be more involved in the discussion and "*give her opinion especially when [participants] are stuck*". Some groups also stated issues with the voice recognition and that sometimes they had to repeat things for the robot to understand. They said they did not find it annoying but time consuming.

5.7 Discussion and Conclusion

Individuals who participated in hiring meetings facilitated by the robot felt positively about the robot and experience, based on self-reported attitudes (Working Alliance, assessments of the robot's knowledgeability and friendliness). Individuals in the ROBOT condition also were also more satisfied with their team and the decision-making activity compared to those who held hiring meetings without the robot (CONTROL condition), and expressed as significantly greater increase in decision confidence in the ROBOT vs. CONTROL condition.

The robot had a definite impact on group behavior, with significant differences in time spent in different meeting activities, adherence to the agenda, and whether they performed certain recommended task steps at all. In post-session interviews, all groups said they would use the robot facilitation system if they could and would recommend it to others. Most (10 of 11) groups volunteered that the robot helped balance participation in the meeting. Participants also felt the robot was effective at keeping the meeting structured and efficient.

Participants did rate the tablet application alone (CONTROL) significantly higher on overall satisfaction compared to the full facilitation system with the robot and tablet (ROBOT). There are several possible reasons for this. First, participants may have had very different frames of reference for these evaluations: tablet applications are likely very familiar to most participants and participants in the CONTROL group may have interpreted the overall satisfaction question as pertaining only to the design of the tablet application GUI, whereas participants in the ROBOT group were focused more on evaluation of their experience with the robot, possibly comparing it to human-human interaction as a point of reference. Second, the robot did have issues with response latency and speech recognition errors, which may have influenced overall satisfaction ratings. Finally, it may be that simply because the robot facilitator was effective at directing group members, they enjoyed the experience less even though it resulted in a more efficient and productive meeting, and I may have seen similar low ratings for a human facilitator.

Limitations

There are several important limitations to my evaluation study, beyond the small number of participants and the biased convenience sample of mostly college students who participated. First, I used the minimal group size of two participants: although the system and architecture are readily scalable, it remains to be seen how the robot facilitator will fare in larger groups. My participants were mostly strangers to each other in an artificially constructed setting, so were likely on their "best behavior". Real meetings, in which participants have histories of conflict and many agendas, are much more likely to have serious imbalances in participation that would be a better test for the robot. My participants were also peers, and my architecture may need to be adapted for meetings in which hierarchy (e.g., the presence of a manager) influences the social dynamics.

Future Work

There are many important directions for future work that are enabled by my system. Handling a wider range of social dynamics, including various kinds of conflict among participants, is an important area of future research. Simply identifying and classifying conflict is an important first step, but the development and evaluation of intervention strategies are equally important. Extending my framework for a wider range of meeting types, with heterogeneous participants (managers, customers, etc.), represents a large number of opportunities to study and facilitate a number of interesting and important social dynamics.

**This work has been published in a long paper titles 'A Multimodal Robot-Driven Meeting Facilitation System for Group Decision-Making Sessions' at ICMI 2019

Chapter 6: DETECTING DISAGREEMENTS IN GROUP CONVERSATION

6.1 Introduction

A group facilitation system requires a set of sensing mechanisms to understand what is happening in a meeting. It needs to see people and where they are gazing at, it needs to listen to people to be able to respond to them, and it needs to make sense of what people are talking about to provide the best interventions. The ultimate goal of the group facilitation system is to provide just-in-time instructions and interventions to improve the decision-making experience. One of the interventions that my group facilitation system seeks to provide is 'disagreement management' during a group decision-making meeting. In order to do so, the system must first be able to identify when a disagreement is happening between two or more people in the meeting. Automated identification of disagreements in a group meeting will enable the facilitation system to provide disagreement management interventions when needed.

In this chapter, I describe a machine learning method to classify utterances as agreement or disagreement automatically. I developed a machine learning model that classifies each utterance as being a disagreement, given the discourse context (prior utterances in the meeting).

6.2 Related Work

Numerous research efforts have leveraged AI techniques to detect specific events such as decision-making moments or disagreements between members in a group meeting. For example, Kim and Rudin applied supervised and unsupervised algorithms on the AMI dataset to predict when important actions related to a decision take place in a meeting [108]. Using SVM-based and Naive-Bayes Gaussian models, they were able to predict key decision points with an F-score greater than 0.8. In the rest of this section, I focus on previous research on automated identification of disagreements and conflicts in spoken conversations using machine learning techniques.

Several studies have used various computational models and classical machine learning models to detect disagreement both in textual data [193] (e.g., Wikipedia talk pages) and in spoken conversations [109]. For instance, Galley et al. used a maximum entropy ranking method to identify adjacency pairs in a conversation first, and then classify utterances as agreement or disagreement using structural, durational, and lexical features. They modeled the discussion context using a Bayesian network that captures the relations and dependencies between utterances [67].

Researchers have used various features sets (e.g., lexical, structural, durational, prosodic, and discourse-related features) related to different aspects of argument expression (e.g., verbal, non-verbal, prosodic, and a combination of them) to train conflict detection models. Germesin and Wilson explored different sets of features that can better detect agreement and disagreement utterances in a meeting [71]. They included lexical, prosodic, and structural features in two machine learning models (a decision tree and a conditional random field (CRF)) and found the CRF model without higher-level information labeling resulted in a more precise prediction of (dis)agreement. Kim et al, compared prosodic and conversational features in detecting different degrees of conflict (low, medium, high) in a political debate. They found that prosodic features outperformed the conversational features in this three-class classification problem [109].

Nonverbal behaviors have also been shown to be a strong indicator of argumentation behaviors in conversations and much prior work has explored and developed computational models to detect disagreement using non-verbal features. Accordingly, several researchers have investigated the automatic detection of (dis)agreement in discourse based on nonverbal behaviors [22]. Bousmalis et al. describe the nonverbal behavior cues of agreement and disagreement in group conversation. They also reviewed several available databases and 8 papers that used machine learning classifiers to detect (dis)agreement in discourse [24].

Using various features enables the models to train on a rich set of relative information; however extracting these features usually requires time and manual effort. Thus, a few papers have investigated semi-supervised machine learning models that are trained on large unlabeled datasets combined with a small labeled dataset [81, 87] to decrease the effort of manual labeling. A variety of machine learning models including decision trees [87], conditional random fields [194], support vector machines [109, 151], Baysian networks [67], and contrast classifiers [81], have been used to detect extreme cases of disagreement or conflicts in political broadcast debates [194], as well as disagreements in group meeting conversations [71].

Although much work has employed machine learning techniques to detect disagreement using lexical, structural, durational, prosodic, and discourse related features, less work has used state of the art computational approaches in Natural Language Processing (NLP) and deep learning models to detect disagreement by considering the meaning of utterances and the conversation context. An advantage of deep learning models is their ability to learn and extract features as a part of the training process. This gives them a great potential to be used in real-time systems. Thus, in my work, I investigated the use of a deep learning approach to detect disagreement in conversations. While the system currently uses the transcripts of verbal communication between participants, the system could extended to using multimodal inputs including the audio (prosodic), as well as emotions (sentiments) of meeting participants in the future. The output from the model can then be used by the intervention module in real-time as the conversation is going on.

6.3 Meeting Corpora

I used two public datasets of small group decision-making conversation; AMI and GAP. Both of these datasets contain text transcripts of group conversation as well as argumentation labels, indicating the argumentation status of each utterance in the conversation. Here I describe each dataset in more detail.

6.3.1 AMI Dataset

The Augmented Multi-party Interaction (AMI) meeting corpus consists of 100 hours of meeting audio and video recordings. AMI is widely used among researchers to study group behavior and develop computational models to predict patterns in a group discussion [71]. The meetings recorded in this corpus are group decision-making sessions with four participants who discuss the design of a new television remote control. Participants play the roles of employees in an electronics company that decides to develop a new type of television remote control. The participants are told they are joining a design team whose task, over a day of individual work and group meetings, is to develop a prototype of the new remote control. In this context, each participant is given a specific role in the meeting. The roles include project manager, user interface designer, industrial designer, and marketing expert.

Annotations: The AMI corpus provides various annotations, including verbal, non-verbal, and contextual labels, along with the transcripts, audio, and video recordings of the meetings. In this project, I only used the portion of the AMI dataset that contains argumentation annotations. The AMI corpus creators used Twente Argument Schema (TAS) [162] to formally annotate the group discussion segments with argumentation labels. TAS is designed to formalize the observed argumentation in a group discussion and captures the important conversational moves in a group decision-making discussion. In order to preserve the conversational flow, TAS defines a tree-based data structure where the speech segments are nodes that can, for example, be strong or weak statements or open issues. In this tree, the edges are relations labels between two nodes (e.g. two statements) that can be, for example, a request, or a positive (agreement) or negative (disagreement) relations.

6.3.2 GAP Dataset

The Group Affect and Performance (GAP) corpus consists of transcripts of group conversations for 13 small group meetings, approximately 104 minutes of audio recording, as well as annotations per utterance. This data is collected and published by Braley and Murray to stimulate the research on computational analysis of small group meetings [180].

In each meeting, two to four participants were involved in a group decision-making task. In total, 37 participants formed 13 groups and attended the sessions. All groups were instructed to perform a ‘winter survival task’ which is a group building exercise similar to the desert-survival task [55]. The winter survival task is a group decision-making activity in which participants are asked to imagine a hypothetical plane crashed situation and need to work together as a team to survive. They are presented with a list of 15 items that remained from the plane and are required to rank these items based on their importance to the team’s survival.

Annotations: Each group meeting was split into speaker intention segments. Each segment is then manually annotated for sentiment and argumentation (decision-making) labels. For sentiment labeling, they used a binary annotation schema with positive and negative sentiment values. Meetings were also annotated for group argumentation statements, which refers to the status of each segment in the group decision-making process. The four argumentation labels are proposal, agreement, disagreement, and confirmation.

6.4 Automatic Detection of Disagreement in Group Conversation

Various modalities can be used to detect disagreement moments in a spoken conversation. In addition to the pure semantics of a sentence that declares disagreement, prior research has shown people’s nonverbal behavior, voice prosody, and even discourse structure, are all affected when they express a disagreement [22, 71]. Much research has investigated the use of various non-verbal, prosodic, and conversational features to train machine learning models for disagreement detection. An important aspect of the disagreement detection model in this context is that it has to provide the prediction in real-time to be used by the intervention module. While I admit the importance of all of these features, as a preliminary step I use only the text of meeting participant utterances to predict its label, as it does not require any further feature extraction. State of the art speech to text technology offers relatively accurate transcripts that can be utilized by the model without further manual manipulation.

To represent the spoken words and sentences in a machine-understandable format, I used word and sentence embedding [133]. Focusing on the disagreement prediction task for each utterance, I evaluated two different language models and prediction techniques and compared the results. I used a sequence-based model (LSTM) and a data-driven transformer model (BERT) to predict disagreements in a multi-party conversation.

LSTM or Long Short Term Memory neural networks are a type of Recurrent Neural Network (RNNs) [132]. LSTM is specifically useful to model and learn order dependencies in sequential data, and they are widely used for speech and language problems [72]. RNNs in general are capable of representing context information by keeping some information about past inputs. However, RNNs fail to keep information from longer time gaps in a sequence because of a computational problem (vanishing problem), and LSTMs were explicitly introduced to solve the long-term dependency problem. Thus, I hypothesized that LSTMs can be useful in detecting disagreement in a conversation by looking at the discourse context (previous utterances) for a conversation.

The Bidirectional Encoder Representation of Transformer (BERT) is a novel approach in language modeling that applies certain key points from previous models: 1) it has bidirectional training, and 2) it uses a transformer architecture and attention mechanism [57]. Using these key features BERT has revolutionized context learning in language models. BERT is being widely used for various NLP tasks such as classification, question answering and Named Entity Recognition. Given BERT's capability in learning the contextual relations between words and sentences in a text, I hypothesize that it would perform well in detecting disagreements in a group conversation. Moreover, BERT can be trained on a large amount of unlabeled data and there are very large pretrained BERT language models that can learn any new language data in an unsupervised manner. In this effort I used pre-trained BERT to transform group member utterances to a fixed-length vector.

6.4.1 Method

Table 7. Distribution of labels in AMI and GAP datasets

Utterance Labels	# of occurrences in our data
Agreement	4267
Elaboration	1619
Disagreement	1099
Option	1341
Proposal	516
Uncertain	514
Confirmation	18

I formulated the automatic detection of intragroup disagreement as a classification problem. My data set consisted of 9,613 utterances, each labeled with one of the 7 classes shown in table 7. I split the data into a training set and a validation set, with 90% for training, and 10% for validation.

Data Pre-processing

The data used consists of 28 transcripts of group meetings from the GAP dataset and 94 meeting transcripts from the AMI dataset. To prepare the data for the deep learning models, all transcripts were first segmented into conversational turns by each speaker. The GAP and AMI dataset together contained approximately 110k utterances. I then needed to find an appropriate numerical representation of words and utterances. To do so, I evaluated multiple word and sentence embedding techniques after initial cleaning of the data (replacing compact form of phrases with full phrases, removing stop words, removing empty utterances). For word embedding I tried two pre-trained models: GloVe on the commonCrawl data set [154], and Word2vec on Google News [133]. To compare the word embedding models, I visualized the numeric representation of words in a 2D environment after dimension reduction (via t-SNE), and I chose word2vec model as the word embedding. I also trained my own word embedding. For sentence embedding, I tested the averaging technique [7] and calculated the average vector of all word vectors in a sentence. The second sentence embedding model I tried was Gensim Doc2Vec model [69]. To evaluate the sentence embedding models I ran cosine similarity and found better results with the Gensim model, so I chose that as my final doc2vec model.

LSTM Model Architecture:

Using an encoder, I mapped the sentence vectors generated by the Gensim model to vectors with size of 500. The vector representation of each utterance besides its 3 previous utterances were fed to the model in each row, resulting in a 3-dimensional input matrix #of utterances * #of sequenced sentences * vector dimension (9613 * 3 * 500). I also prepared a one-hot-vector representation for each of the 7 labels. I then fed the input matrix to a Bidirectional LSTM model. The Bidirectional wrapper is used with a LSTM layer. This helps LSTM to learn long term dependencies. I then fit it to a dense neural network to do classification. I used a relu function and finally added a Dropout and a Dense layer with 7 units and SoftMax activation. When I have multiple outputs, SoftMax converts outputs layers into a probability distribution.

BERT model Architecture:

The second model I evaluated was BERT. For the BERT model I simplified the problem to a binary classification. This model is only designed to predict whether an utterance is disagreement or not. I loaded the cleaned and time- ordered utterances from the AMI and GAP datasets. To balance the distribution of classes I resampled the utterances with disagreement label, resulting in 5,500 disagreement utterances out of a total of 14,014 utterances. I used the BERT tokenizer to get fixed-size (32) vector representations of each utterance. I then fine-tuned the pre-trained the BERT model by passing utterance vectors to it.

I split my data into training and test datasets, and built a custom model containing BERT, with a dense layer and a softmax layer at the end. I performed the classification task for each utterance using the CLS embedding induced by BERT.

6.5 Results

I conducted experiments with both datasets described in Section 6.3. Table 8 shows the performance of the LSTM and BERT models. Among LSTM models, I got the best performance when considering 5 previouse utterances (69% acc, 69% F1). A binary classifier using BERT had the highest accuracy on the test dataset (78% acc, 71% F1). Considering more labels resulted in lower performance of the BERT model. I tried three labels: Agreement, Disagreement and Elaboration, and the accuracy decreased to 69%. The use of more labels for this problem is not required, given that the goal is to detect only disagreement utterances.

Table 8. Accuracy and F1-scores of different models on the training and testing datasets

Model	Training Accuracy	Testing Accuracy	F1-score	# of labels

LSTM-3	0.88	0.64	0.63	7
LSTM-4	89	68	67	7
LSTM-5	**0.92**	**0.69**	**0.69**	**7**
LSTM-6	0.82	0.66	0.65	7
BERT	0.76	0.69	0.68	3
BERT	**0.91**	**0.78**	**0.71**	2

6.6 Discussion and Design Implications

Although I have not evaluated my models in real time, by developing these models I showed the feasibility of using deep learning models with low real-time computational demands for disagreement detection. My models can be certainly improved by adding more features and doing more contextual analysis such as speaker-addressee detection. Three natural next steps for my disagreement detection model are:

1) Include other features such as audio and sentiment which are known as good indicators of disagreement yet can be automatically extracted.
2) Use a longer history of utterances to detect the argumentation status of the current utterance.
3) Use the previous utterances of the same speaker to better understand the context to detect disagreement.

Another interesting future step is to add a learning capability to the model. Getting feedback from previously detected labels in the same meeting and including the feedback to re-train the model could be very effective in learning the specific features of a meeting, and language of its members.

6.6.1 Disagreement Detection Models Evaluation

A machine learning model should be evaluated in context, and based on the consequences of its prediction outcomes. Considering the context of group facilitation, as well as the aftereffects of a false positive and a false negative prediction, here I discuss the interplay between the detection and intervention module to optimize the performance of the group facilitation system.

Ultimately the output from such a model will be linked to an intervention module, which controls the robot to deliver disagreement management instruction. Therefore, errors of the detection module will have minor or major implications in the overall performance of the system. Two scenarios could follow a false detection.

The first scenario is when the model detects an utterance as disagreement, but actually, there was no disagreement (false positive). Such error could result in interruption and break the flow of participants' conversation. Without a repair strategy (e.g., the robot continues the disagreement management without confirming if the case was disagreement), such an error has a high cost for the group, as it can derail the whole conversation and decrease the efficiency. This conveys the importance of seeking confirmation by the robot, as reflected in the qualitative findings in section 7.6.

The second scenario happens when the model recognizes no difference between opinions (does not detect a disagreement), while participants have actually opposing ideas (false negative). The consequences of such errors depend on the level of the necessity of intervention in a given situation. If participants expect the robot to intervene, and their next move is dependent on the robot's mediation, then a false negative is also costly for the group. Thus, it is crucial to design the whole interaction and intervention to compensate for the potential errors in the detection module. For example, the robot should explicitly ask for each participant's opinion at important decision points when the robot's intervention is needed in case of a disagreement. Moreover, the robot should leave room for flexible conversations and be programmed to receive urgent messages. For instance, when the model does not detect a disagreement and people want the robot to do something about it, they may just call the robot and say a command to show a disagreement has arisen.

In sum, I believe in the context of my research, there is a strong tie between disagreement detection and intervention modules in a way that the performance of the detection model could be compensated for or worsened by the interventions delivered by the robot. This is why I moved on to investigate how a virtual group facilitator can provide interventions for disagreement management in a group decision-making scenario.

Chapter 7: DISAGREEMENT MANAGEMENT INTERVENTIONS FOR GROUP FACILITATION ROBOTS

In Chapter 4, I showed the feasibility of using an embodied conversational agent for group meeting facilitation. Chapter 5 then reported the design and development of an *automated* group facilitation robot to help small groups in a group decision-making scenario by providing several services such as; managing the meeting structure, managing the time, and balancing group member participation. In this chapter, I specifically focus on group conflicts (a.k.a. disagreements) as a very common challenge in group meetings that could negatively affect group performance and satisfaction. I examine how a robot facilitator can provide real-time interventions to manage group conflicts in decision-making meetings. I improved the meeting facilitation services (structure, time and participation management) of the group facilitation robot based on findings of my previous studies, and incorporated a new decision-making facilitation module. Decision-Making facilitation includes: 1) providing intelligent interventions for disagreement management (either ACTIVE or PASSIVE strategy), and 2) offering information related to the decision-making task. To evaluate this approach, I conducted a between-subject experiment with 26 participants (forming 13 groups) who interacted with a semi-automated robot. The study compared two strategies of disagreement management delivered by the robot facilitator: ACTIVE disagreement management, and PASSIVE disagreement management. In this chapter, I report the quantitative findings of the between subject user study.

7.1 Introduction

Intragroup conflicts are unavoidable when groups of people work together [76]. Conflict or disagreement arises when interactants oppose each other's views and express diverse opinions. Conflict among members of a group can explain important variance in the group outcome and performance [75]. A recent survey estimated that in the US alone, 385 million working days are spent every year dealing with conflict in the workplace [84]. Moreover, intragroup conflicts have a dynamic nature [75]. Depending on its type (e.g., relationship, task, or process conflict), and when it happens in the meeting life cycle, conflict may have different impacts on the group. Numerous researchers have studied intragroup conflict and demonstrated both positive and negative effects on group performance. Although empirical evidence about the effects of different types of conflict on group performance is mixed, researchers agree that leaving a conflict unmanaged can be destructive to any meeting [76]. To avoid detrimental effects on individuals as well as group functioning, groups should identify the causes of conflict and take appropriate action to manage them [12]. Conflict management is essential for confining the negative effects of conflicts [12], and much research has explored ways to manage conflicts in small groups.

The ways to manage a conflict situation may be as varied as to its causes and contexts[12]. Numerous strategies and techniques were introduced for conflict management, aiming to minimize the disruption caused by disagreement and provide a solution that satisfies all sides of the argument. These strategies are either undertaken by the parties in an argument or involve interventions of an outside party [12]. The problem-solving or collaboration approach is a conflict management strategy that "attempts to achieve close collaboration and integrative decision-making between individuals" [12]. Although collaborative problem-solving has been shown as the most effective method for dealing with an underlying conflict [29, 117], it has requirements that may be absent in a dyadic conflict situation. Requirements such as flexibility of interactants, power symmetry, trust, and confidence in each other, lack of biases, and lack of negative impulses towards each other may not exist among parties of an argument [12]. As a result, in certain cases, individual's responses in a dyadic conflict may further escalate the conflict rather than resolve it. A potential solution to this is that a third person intervenes to facilitate the problem-solving conversation by creating conditions which are suitable for conflict resolution [12, 144, 165]. I call this third-party conflict management.

In my research, I suggest using a social robot as a third-party to attempt to resolve a conflict or disagreement among members of a group. Kats and McNulty identified reflective listening and maintaining a positive emotional atmosphere as keys to effective conflict resolution [142]. Inspired by this work and various third-party conflict management methods [123], I propose a step-by-step conflict management procedure to be delivered by the group facilitation robot when a conflict arises in the group. The facilitation robot will have an explicit verbal and procedural intervention to resolve the disagreement. I adapted a step-by-step approach suggested by Michael Wilkinson[1] to handle task disagreements in the group. The robot takes the following steps when a disagreement is detected: The robot 1) starts with declaring **agreements** to establish a common ground between the parties. 2) practices **active listening** to confirm and **clarify** the source of the disagreement. 3) Asks the parties to discuss the **pros and cons** of their unfavored option. 4) Asks each party to review the **alternative** options (asks **delineating questions** such as how much does it cost, what is involved, etc.). 5) **Summarizes** the discussion. 6) Checks with the parties to determine whether consensus has been reached.

After showing the overall feasibility and acceptance of a group facilitation robot, I wanted to test the performance of such robot in mediating specific challenges in a group meeting, such as disagreement/conflict situations. In this chapter, my goal is to explore how a group facilitation robot can play a role in managing group conflict. To achieve this goal, I designed and developed a group facilitation robot with disagreement management interventions as a decision-making facilitation. I evaluated the robot, as well as its disagreement management, and meeting facilitation interventions, by conducting a Wizard-of-Oz style, in-lab user study. I conducted a between-subject experiment, with 26 participants who attended group decision-making activities facilitated by the robot. In order to evaluate the effects of the conflict management interventions by the robot besides understanding overall perception of the facilitation robot, half of the participants experienced the robot with limited conflict management interventions (PASSIVE condition), while the other half interacted with a robot who provided more conflict management strategies (ACTIVE condition). Through survey responses, I first examined the effects of different types of conflict management interventions that I report in this chapter. Then, to find out about the unanswered questions from the subjective measures, I conducted semi-structured interviews with all participants to learn about their perceptions of and expectations from the robot. I also sought to understand the perspective of experienced information workers who do attend group meetings regularly together and examine how they perceive the robot in real-world group setting. Thus, I conducted a focus group with members of an existing group at the administration department of an educational institute. I describe the design of the focus group and the findings of my qualitative analysis in the next chapter.

The driving research questions for this work are RQ5) How can a social robot facilitate a small group decision making session and resolve specific challenges such as intragroup conflict? What is the participants' attitudes towards the group facilitation robot, and towards how the robot can help in conflict management? And RQ1) What would information workers expect and want from a social robot in a real work setting?

In the remainder of this chapter, I first review related work that motivates my study. I then present how I designed the group facilitation agent, the semi-automated robot ("Wizard of Oz") experiment I conducted, as well as the experimental methodology. In the results section I examine participants' subjective perceptions from survey responses. As mentioned above, in the next chapter, I report the extensive analysis of the interviews and focus group data to inform the future design of social robots for group meeting facilitation.

7.2 Related Work

I reviewed most of the prior work related to conflict management and the use of robots in groups in Chapter 2. Here I briefly review the research in Human Robot Interaction (HRI) that has explored using robots in groups to help in performing tasks, facilitate a group activity, or moderate a team conflict.

7.2.1 Robots in Groups and Teams

The history of using robots as collaborators in groups can be traced back to early 2000 when Hinds et al. studied the effects of appearance (human-likeness) and relative status (subordinate, peer, and supervisor) of the coworker robot on user satisfaction and collaboration outcome. Prior to that, though, Nass et al. had started exploring the use of computers at teammates [141]. Leveraging the advances in sensing and speech technology, researchers developed more complex robots to collaborate with humans in a range of situations, from small teamwork [175]to high-stakes teamwork such as rescue missions [140], and invasive surgeries [60]. Much of the research on robots at groups have been focused on one-on-one collaboration between a robot and a human in a group setting or at the workplace. Fewer studies, however, explored the interaction of a robot with multiple humans in a group. A few examples are studies conducted by Malte Jung and his group. In [153], they studied the effects of the robot's verbal and nonverbal behavior on conformity in small group decision-making sessions both on conscious and unconscious levels. They also extensively studied the socioemotional effects of a robot on group members and proposed 'affective grounding' as essential to coordination for both individual and group interaction between humans and robots [99]. Similarly, Correia et al. have conducted multiple studies exploring trust [45], emotions [46], and membership preferences [47], in game teams of human and robot. Most of prior research has examined the use of robots in groups to perform or assist humans in performing some tasks (e.g., rescue mission[140], surgery [60], delivery [163], manufacturing), however introduction of social and semi-intelligent robots suggests that the involvement of robots can go beyond performing task. On group level, robots can affect social interaction, change people's communication, and evoke emotion [102]. Jung also discussed how robots affect collective human interactions by a group's social functioning, besides the task-specific assistance that they provide [102]. Next I review the existing yet thin body of research on social and facilitative roles of robots in groups.

7.2.2 Robots Facilitating Groups

With advances in sensing as well as natural language understanding technology, multiple research groups started delegating higher-level responsibilities (e.g., team moderation) to the robots in groups. Matsuyama et al. presented a framework for a robot that can control and regulate imbalance engagement in a small group conversation. They used a Partially Observable Markov Decision Process to model how a robot should *observe* the conservation status, *obtain an initiative* and take some action towards the *floor and topic management* [127]. In their proposed framework, the robot first detects when a participant left behind a conversation between two other participants (A and B). The robot then joins the dominant discussion between A and B, and try to get the floor like a harmonized agent. Finally, the robot initiates a question to involve the left-behind participant in the conversation. Matsuyama *et al.* have evaluated the effects of their proposed system on the feeling of groupness and perception of the robot's timing and initiating process by having participants watch 30-second videos of 4-member group conversations. Their results indicated that the procedural approach to managing the floor resulted in higher acceptability and a feeling of groupness. Although their evaluation was via acted videos, not through open interaction with the robot, their analysis showed the best timing for initiating a procedure is after the second or third pair of turns versus after the first adjacency pair [124]. Socially Assistive Robotics (SAR) is also a subfield in HRI literature focusing on developing computational models that enable robots to assist users in goal-directed interactions (e.g., in health and education) [168]. Short and Mataric incorporated SAR methods into multiparty interactions, and presented an algorithm to provide task moderation and social moderation in a team playing a game. They developed a small dragon-like robot that uses their proposed algorithm to assign conversational and physical resources in a game with the goal of moderating the team [168].

MicBot [179], is another research effort in which the authors examined the use of a peripheral robotic object to promote participant engagement in a small group problem-solving setting. They developed a robot in the shape of a microphone that can show certain non-verbal behavior (movements to show follow or encourage actions) to promote group engagement and performance.

A few research studies in HRI particularly investigated robots for conflict management in groups. Jung et al., explored groups' perception of a robot that attempts to repair interpersonal violations among team members during a problem-solving group task [101]. In a Wizard of Oz experimental setting, they evaluated conflict moderation interventions delivered by a physically and verbally interactive robot collaborating with a team of 3 individuals (2 participants and a confederate) on a bomb defusal task. The confederate triggered either personal or task-directed negative attacks to participants, and the robot responded by either positive repair or neutral approach. Results of a between-subject experiment on 53 teams indicated that the robot's repair is only effective for personal attacks. Group performance was not impacted by the robot, however, teams perceived more conflicts when the robot attempted to repair a negative personal attack compared to when the robot provided neutral comment. Shen and Jung also explored a conflict resolution robot for children with the goal of teaching conflict resolution skills to kids at an early age. To evaluate the robot, they compared a control condition (robot only directed the play session with no conflict intervention) with a conflict moderation condition in which the robot offered step-by-step prompts for constructive conflict resolution. They found that 'children were more likely to move on and resolve conflicts constructively in the mediation condition than in the control condition.'[167]. Unlike these studies, I seek to investigate conflict management in a group decision-making setting, using a procedural conflict resolution intervention that is inspired by the literature in conflict management at the workplace. My robot's design and interventions target adults and professionals who collaborate at teams. In a formal setting such as a workplace, a facilitator with a human-like look and behavior has great promise as an effective medium for delivering conflict resolution interventions. Last but not least, the interventions in my system do not limit to mediations after the conflict onset, but it contains a series of meeting facilitation and decision-making facilitation (e.g., structure and participation management) to prevent unnecessary disagreements by clarifying the arguments.

In a different realm, researchers investigated the processes involved in establishing and maintaining trust in human-robot teams. In [173], the authors examined human teammates' perception of social robots with two failure repair behaviors: 1) a robot that makes neutral statements after making mistakes (ignore and no admit), 2) a robot that makes vulnerable comments and admit to any mistakes made at each round of a collaborative game. Their results demonstrated that robot's vulnerable statements in admitting to mistakes could have affect teammates' trust negatively.

7.3 SoFi: A Social Robot for Group Meeting Facilitation

This section outlines design considerations and features I defined for the group facilitation robot with disagreement management interventions in the third prototype. This robot's design is mainly inspired by the literature in the management sciences on the role of a group facilitator, as well as the previous work on the group decision-making support tools (including my own previous studies reported in chapter 4 and chapter 5). In this chapter, I introduce SoFi, a social robot for group facilitation that can communicate with group members via conversation, and provide disagreement management interventions and informational support on top of social and meeting facilitation services. The long-term goal of this group facilitation system is to improve the group decision-making experience for the members of the group.

SoFi is a social robot that facilitates group decision-making meetings. SoFi is designed to support collaborative activities in a group meeting by providing instructions, information, and interventions to alleviate some of the known challenges in a group meeting, such as deviation from the agenda, dominance, and conflict. SoFi has a human-like head and face that enables it to gaze different directions, and show facial expression [Figure 10]. Furthermore, to facilitate natural interaction with a group, SoFi is able to conduct quasi-natural conversations with all group members. It also shows attentive listening behaviors, allowing group members to engage in a conversation with it.

SoFi is represented with a humanoid head developed by Furhat robotics[9] [Figure 10]. In this study, I used the new version of Furhat that has built-in microphones and camera. Similar to the robot I described in Chapter 5, this robotic head has an animated face, back-projected on a translucent mask, mounted on a mechanical neck that allows the head to turn and nod [3]. In addition, the new version has a three degree of freedom that allows the head to turn, node, and roll. Adding the third

[9] https://furhatrobotics.com/

degree of freedom has made the robot's head movements and gazes look much more natural. The robot's speech is generated using the Cereproc speech synthesizer. The robot can direct its attention using eye gaze and head pose, show facial expression, and listening behaviors by moving her eyebrows, lips, nodding and rolling her head.

7.3.1 Design Requirements

My first study has demonstrated the benefits of using and embodied conversational agent as a group facilitator over a voice-only agent [Chapter 4]. Previous research has demonstrated the superiority of physical robots to screen-based agents in engaging users in a multiparty conversation [3, 138]. Robot's physicality also improves users' perception of its presence in group interaction [36]. A physical robot can use its head movement and gaze direction to more accurately display attentiveness and feedback to speakers and to indicate the desired addressee [120]. Thus, I chose to use a physical robot as the group facilitator [Figure 10].

Figure 10. The study setting where SoFi is guiding and facilitating the group decision-making session with two participants

The main goals of the system are to support different needs of a group during a decision-making task, facilitate the group task, and improve overall meeting quality by providing just-in-time interventions. Inspired by the literature in business and organizational sciences the robot is designed with specific capabilities to mitigate the common challenges in group meetings including: diverging from the agenda, discussion domination, conflict, and time inefficiency[148, 150]. I categorize the main facilitation functions of the robot under three different types of support during a group decision-making session: social facilitation, meeting facilitation, and decision-making facilitation. Here, I briefly review each facilitation type.

Social Facilitation

Catalyzing the social interaction among members of a workgroup is one of the most important tasks of a group facilitator [85]. A group facilitator usually starts a session and welcome and thank everyone for their attendance. He/she starts to build a trustworthy relationship with group members to improve a sense of teamwork [56]. Thus, to simulate the social behavior of a human facilitator, the robot facilitator needs display facial expressions, as well as listening and understanding behaviors. It needs to be able to create a safe and comfortable climate for the members to express their thought. It should also build rapport, and show empathy by engaging group members in natural, multimodal, face-to-face conversation, to be accepted in this social role.

Meeting Facilitation

Group facilitators provide real-time interventions and instructions to alleviate common challenges of group decision-making meetings, such as inefficient and unstructured discussion, and imbalanced participation (dominance) [189]. Likewise, the group facilitation robot needs to be able to enforce a structure to the group conversation, monitor members' participation and actively nudge less active members to balance the participation, and finally manage the time of the meeting.

Moreover, a robot that takes the role of a group facilitator should be perceived as an authorized and intelligent member to enforce participants' attendance and adherence to its instructions. Therefore, as suggested by [34], a group facilitation robot should be able to simulate human-like verbal and nonverbal behaviors such as gaze and facial expression to allows participants to treat it as an intelligent and authorized social entity with a persona.

Decision-Making Facilitation

In addition to the general meeting facilitation functions, a group facilitator requires to support the group decision-making activities in particular. To help a group make a decision more efficiently a group facilitator should be able to provide background information (either directly or by querying other resources) related to the task being discussed.

Furthermore, offering appropriate interventions to manage disagreements amongst members of the group, and converging different opinions on the same task, is a very critical yet challenging task for group facilitators [123, 135]. Inspired by the human resource management literature [12, 200], I propose that a robot facilitator's conflict management interventions can be categorized into *prevention* tasks and *management* tasks. Conflict prevention tasks are the ones that are recommended in any group discussion to assure everyone has a chance to express their thoughts. To *prevent* a misunderstanding or disagreement the facilitator enforces a structure allowing the members to listen to each other and come to a conclusion before they jump to the next topic. While, conflict *management* tasks include the actions that are taken after a conflict or disagreement arises among two or more members of the group. A group facilitator should take multiple steps upon the detection of a disagreement in the group discussion to resolve the conflict. For example, a facilitator should first attempt to clear the discussion ground and then take procedural steps [12] to make both sides of an argument reach to a common understanding or a solution.

Multiparty Conversation Competence

The last design requirement reflects the importance of a group facilitator's competence in handling a multiparty conversation. Involvement of multiple humans in the interaction, is an important feature of group decision-making session, and a group facilitator always pays attention to engage all members in the group conversation [56]. Thus, a robot facilitator should be able to support multi-party conversations well. To conduct multiparty interaction with all group members, the robot needs to receive speech and gaze inputs from each member. The robot also must be capable of performing minimal turn-taking management, indicating attentive behavior by gazing at the speaker, as well as indicating her intention for a specific user to speak by gazing toward them.

7.3.2 SoFi's Facilitation Features

Based on the design guidelines described in section 7.3.1, here I explain the specific features and functions I designed for the group facilitation robot (SoFi) to address the requirements and provide different types of facilitation.

Social Facilitation Functions

To establish a comfortable and safe environment for the group discussion SoFi initiates the meeting by welcoming the team, greeting everyone and introducing herself and the members of the group. She introduces herself as a facilitator who is with the team to support their decision-making. She calls each person by their name and asks them to talk about favorite part of their work. This serves as "ice-breakers" to acquaint group members with each other and build trust and rapport. She then sets the meeting stage by orienting group members to the meeting goal, task and process; discussing desired meeting outcomes; and reviewing the agenda. In order to build rapport and trust as a social facilitator, SoFi displays understanding and attentive listening by directing her gaze, smiling and doing head nod, head roll and head shake.

Meeting and Decision-Making Facilitation Functions

SoFi is designed with specific capabilities enabling her to provide meeting and decision-making facilitation to a group.

• Enforce meeting structure: To avoid deviation from the meeting agenda, SoFi enforces the meeting structure by offering step-by-step instructions.

• Encourage and Balance Participation: To ensure balanced participation, and to create a safe space for all participants to express their thoughts, at every decision point SoFi asks each member what they think. If the robot notices some participants are significantly less vocal during a meeting, she engages them in the conversation by asking them whether they want to add anything else.

• Manage the Time: During the decision-making meeting, SoFi announces how much time ideally should be spent on each task. She keeps track of the time spent on each stage and remind the group of the time on demand. She may also ask the group to move forward if they stuck at a decision for more than 2 minutes. The system includes a timer display to remind the participants of the remained time [Figure 10].

• Manage Intra-group Conflicts: Motivated by the conflict management strategies in the group behavior literature, I designed a procedural conflict management intervention to be delivered by SoFi when a conflict arises in the team. SoFi's intervention includes 5 steps that take place one after another. When group members have opposing opinions about a decision the robot 1) first acknowledges that there is a disagreement and invite the team to resolve it in a collaborative manner, 2) then SoFi asks the group to practice "active listening" and guides them through the practice step by step. 3) If they still disagree, SoFi asks each member to review the pros and cons of the options being discussed, and again check with each member whether they are convinced to change their decision. 4) If they still disagree, SoFi asks each person to review the alternatives to their proposed option, reassess what each member thinks about the decision. 5) Finally, SoFi suggests the group to finalize their decision and summarizes the final decision (either agreement or disagreement). More details on the disagreement management interventions and SoFi sample dialogue can be found in section 7.4.2.

• Provide background information: The robot has access to extra information about the task being discussed in the group. This information can be stored locally or by querying to other online or offline resources. SoFi provides this information as needed when she is guiding the discussion and is ready to provide more when she receives a request from members of the group.

7.4 Study Design - Methodology

To answer my research questions, I designed and developed a conversational humanoid robot that can facilitate a small group decision making task. I first conducted a controlled user study to assess the user's attitude towards the facilitation system and disagreement management interventions. In this between-subject study all participants experienced the main facilitation function of the robot, while I compared different conflict management strategies by having half of the groups experience a more active conflict management intervention enforced by the robot. I chose an adapted version of the Winter Survival exercise to be the task of the group decision making session [section 7.4.1].

7.4.1 Task: Winter Survival Task

I used a modified version of the standard winter survival exercise as the group task in this study.

Although the proposed design for the group facilitation robot can be used for most types of group decision-making sessions, I chose a specific group task to investigate the right design for the group facilitation robot in an experimental study. Winter survival task is team building activity that has been studied for group behavior analysis [180]. This task involves several decision-making activities and does not require any expertise or deep knowledge to decide about the items. I modified the task to make it more structured and limit the potential variations in the decision-making and group discussion. Here I describe the modified version of the winter survival task:

Participants are asked to imagine themselves in a hypothetical survival situation in which their plane crash landed in very cold weather and they must collaborate as a team and make multiple group decisions to survive in a plane crashed situation. They are given a list of 10 items and need to decide about the importance of these items in their team survival. A task description is provided and the facilitation robot guides the session and provides instructions.

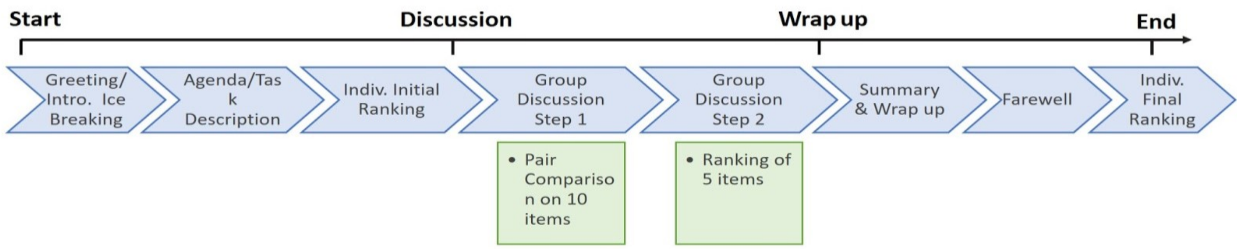


Figure 11. Group decision-making lifecycle

During this session participants first rank the items individually based on the items' importance to their survival. Then they are asked to review and discuss the items in their group to make group decisions collaboratively and the robot supports their meeting by providing instructions and guidance besides specific meeting and decision-making facilitation. The group discussion has two steps:

Step1) Pair-wise comparison of items: The robot initiates this step by asking one of the participants to pick an item to discuss. When a participant picks item A to discuss, the robot provides basic information about item A, and then asks the team to compare item A with item B (after providing information about item B) in terms of their importance to the team's survival. The robot guides the group through the same process 5 times to discuss all 10 items. The paired items and an example of the robot and group dialogue are shown in Table 9 and 10.

Step2) Ranking of the remaining items: In the second step the robot takes 5 items and asks participants to discuss and rank the items based on their importance to the team's survival. To keep the second step as similar as possible for all groups, the robot starts this step with a fixed set of 5 items based on the standard solution for the winter survival task (regardless of what each team decided in the first step). The list of items in as well as an example of the robot and the robot dialogue are shown in Table 9 and 10.

Table 9. List of items to be discussed in the first and second group discussion

Step1 – Pair of Items	Step2 – Items
• Compress kit vs. Extra shirt and pants	• Extra shirt and pants
• Ball of steel wool vs. 30 feet of rope	• Cigarette lighter without the fluid
• Cigarette lighter without the fluid vs. Flash light	• Flash light
• Loaded .45-caliber pistol vs. Pocket Mirror	• Pocket Mirror
• Family-sized chocolate bar vs. Whiskey	• Family-sized chocolate bar

After the group discussion they will be asked to indicate their decisions again individually.

7.4.2 Disagreement Management Interventions

During the group discussion phase the robot provides conflict prevention and conflict management intervention to help the group have a smooth decision-making process. The conflict prevention tasks include managing the structure of the meeting by guiding the group to follow the agenda and balancing participation to assure everyone's voice is heard in the discussion. Here I describe the active and passive conflict management strategies in more detail.

ACTIVE Disagreement Management

The group facilitation robot takes up to 5 steps upon the detection of a disagreement in the active conflict management strategy. 1) Acknowledges the disagreement, and highlights the points of agreement. 2) practices active listening, 3) asks the team to review pros and cons of options, 4) asks the team to review the alternative options, 5) finalizes the decision and report the summary. Table 10 shows SoFi's sample dialogue delivering these 5 steps. Here I describe the 5 steps in more detail:

1) Acknowledge the Disagreement and Articulate the Emerging Consensus: This action creates a common ground and allows members to be clear on the decision in front of them. Moreover, reviewing the emerging consensus amplifies the positive impulse and boosts the discussion climate.

2) Active Listening: Prior to starting the group discussion, SoFi describes to the team what active listening is and how practicing it can improve group communications. Then during the discussion when two members have opposing opinions about an issue, the robot asks them to practice active listening. SoFi first asks member A to describe what s/he thinks and asks member B to carefully listen and then repeat what person A said in her own words.

Active listening is a communication skill that requires the listener to thoroughly attend to what is being said. Active listening is widely used to improve various types of dyadic communications including marital interaction [44], parent-youth relationship [79], and doctor-patient interaction [63], and dispute negotiation [145]. Practicing attentive listening has been shown to derive more agreements in negotiations [64]. Moreover, prior research has revealed positive effects of active listening on parties' satisfaction and rapport in dispute and negotiation conversations [156].

3) Review Pros and Cons: After finishing the active listening process, SoFi asks each party their opinion and if they still disagree, SoFi moves to the next disagreement management intervention: reviewing pros and cons. SoFi, asks each involved party to review the pros and cons of their unfavored option. This instruction enables them to actively seek pros for their disliked option.

4) Review Alternative Options: In case the group does not reach to a consensus after step 2 and 3, the robot requests each participant to suggest an alternative option.

5) Summarize and Wrap up: Finally, SoFi again asks each member about their decisions. If they agree on an option, SoFi announces the final decision, thanks them, and moves on to the next step. If they still disagree after three rounds of interventions, SoFi tells them that it is fine and moves on to the next task. In step 2, when the group has to decide ranking of items, in case of disagreement, SoFi lets the group know that to make the task easier in the next round, she uses a random number generator to randomly pick one of the two items as the n^{th} item in the list. SoFi also insists that this won't be counted as points for any of the participants. Table 8 shows sample scripts for each of these steps.

Table 10. The disagreement management procedure along with sample dialogue delivered by SoFi in each step different steps

Step#	Agent's action	Sample script
1)State agreement	Acknowledges the disagreement and point to an emerging consensus	SoFi: "I think you have different opinions here. which is totally fine. let's see if we can work together to build a solution that would address everyone's concerns. Well, I think you both agree that you need a tool for signaling."[the common benefit of mirror and pistol]
2) Active Listening	Describes Active Listening before the group discussion	SoFi: "Listening is an integral part of effective communication. Most people hear their partners rather than listen to them. These are different from each other. Hearing is a passive physical activity involving the reception of sound. whereas listening is a conscious and deliberate act, that converts the sound signals, into something more meaningful and important. During this task today, I will ask you to practice this skill."
	Asks the team to practice active listening when a they have a disagreement	SoFi: "Now I would like you to practice the active listening skill when your teammates are describing their opinions."
	Guiding the team through each step of active listening: Asks user A to express their thoughts. Describes to user B to listen and reflect	SoFi gazes at user A. SoFi: "well, [user A's name], tell [user B's name] what you think about [the item under discussion], and [the paired item] SoFi: [user A's name], I want you to tell [user B's name] what you think about these items. SoFi gazes at user B. SoFi: "and [user B's name], I want you to listen to what [user A's name] says, and repeat back to him/her, what you heard, in your own words." Repeats for B
	Asks user B to repeat when user A is done:	SoFi: "Well, [user B's name], could you please repeat what [user A's name] just said."
	Thank user B for listening and repeating	SoFi: "Thank you, [user B's name]"
3)Pros & Cons	Asks each participant to review pros and cons of their disliked item	SoFi: "that is Fine. now, I would ask you to talk about the pros and cons of each item regarding the situation we are stuck at." SoFi gazes at B SoFi: "could you please tell us more about the pros and cons of [A's selected item]?" Repeats for A
4)Alternat ives	Asks each participant to discuss alternative options	SoFi: "Well, let's dive a little deeper and discuss the alternative options. Gazes at user A SoFi: Tell us about what can be an alternative to these advantages of [A's selected item] Repeats for B Remind them attentive listening
5)Wrap up	Wraps up, and provides summary	Sofi: Well. so we agree that [item x] and [item y] are useful for signaling in a situation like this. SoFi: "considering what have been discussed which item would you pick now?" [SoFi announces agreement or disagreement and moves on]

PASSIVE Disagreement Management:

In the control condition the robot provides the same conflict prevention tasks (structure and balanced participation), however it only offers passive disagreement management interventions. In this condition, when a disagreement arises, the robot only 1) acknowledges the disagreement, and 2) asks the group to resolve it, by saying "Well it seems you have different ideas about this item, why don't you try to work it out?"

7.4.3 User-Study Design and Procedure

I adopted a between-subjects design to compare the effects of active and passive disagreement management style on the groups. Each experiment session was randomly assigned to one of the two conditions. Upon users arrival, a research assistant described the session procedure and the decision-making task, obtained participants' informed consent. Afterward, participants were asked to fill out a set of questionnaires including demographics, their attitude towards the group and PANAS (Positive And Negative Affect Schedule), Participants were told how much time they had for the task and were guided to the room with the facilitation robot to start the experiment. In the experiment room, the research assistant first introduced the robot and explained how she will see and listen to the participant. The research assistant also explains about the shared display, the timer, and how participants should use the tablet for initial and final ranking of their decisions. The research assistant left the room after starting the program. The research assistant monitored and controlled the robot from another room and came back to the study room when the session ends. After attending the group decision-making session facilitated by the robot, participants were asked to move to different rooms to fill out post-session questionnaires. In order to gain a better insight on participants' attitudes toward my system, I also conducted a semi-structured interview with each group at the end. Finally, participants were given a debriefing document disclosing that a human was controlling the robot (Wizard of Oz), the study goal, and the other study condition. The study was approved by the Institutional Review Board at our University, the whole session took about 90 minutes, and participants were compensated for their time.

Participants: 26 participants were recruited by posting online advertisements in online recruiting websites and posting flyers in the university campus. Participants were required to be at least 21 years old and have had some sort of professional experiences to be eligible for the study. Participants consisted of a mix of students (92%) and full-time employees. They were randomly matched together to form 2-member groups. The average age of participants was 24.3(2.6) and 46% of the participants were male. To minimize the effect of gender and expertise on the group performance, I ensured that the ratio of mixed-gender/same-gender groups, was equal in the two conditions. I avoided grouping people from the same department/program, or different expertise levels together. 90% of participants did not know each other before the sessions.

Measures: To address my hypotheses, I measured several subjective responses from self-report questionnaires. 1) Participants' affective state was assessed through PANAS scale[195] and SAM scale [27] before and after the decision-making session. 2) To answer my research questions, I adapted standard scales [101] to assess participants' experience in terms of their disagreements and conflicts. I also asked them about the disagreement management interventions by the robot. 3) Participants' attitudes towards the robot's usability and personality were assessed using single item questions about the robot (e.g. trust, helpfulness, the robot's knowledgeability, authority, and friendliness). I also assessed users' perception of rapport with the robot via a standard scale used in previous studies (Chapter 4). 4) Users' perception of their team cohesion and their overall satisfaction with their decision-making process were estimated via composite measures. 4) Participants' perception of their meeting performance and efficiency were measured via single item questions. 5) I adapted a standard scale to measure the active listening skills and behaviors (ALAS) [113].

7.5 Quantitative Results

Across both conditions, participants were fairly satisfied with their group decision-making experience (Mean = 6.5 (SD = .55)). They were satisfied with the robot (Mean = 6.23 (SD = 1)), and found it helpful for their decision-making (Mean = 5.92 (SD = 1.3)) (both measured via a single item), and rated its facilitation functions significantly higher than neutral ($t(25) = 10$, $p < .0001$) (measured in a composite scale). They also rated their team cohesion (Mean = 6.48 (SD = .67)), and meeting experience (Mean = 6.33 (SD = .79)) significantly higher than neutral in both conditions ($t(25) = 14.7$, $p < .0001$, $t(25) = 11.4$, $p < .0001$) with no significant differences between conditions. Moreover, on average all groups showed increased positive affect and decreased negative affect after the session. I found no significant differences between conditions regarding team evaluation, task completion time, and change and quality (compared to expert ranking) of their ranking.

To assess the disagreement management interventions, I analyzed the video recording and found that groups in the ACTIVE condition followed 100% of the robot's instructions in disagreement management instructions such as practicing active listening, and reviewing pros and cons.

I did not find any significant differences between the two groups in terms of their initial and final ranking. I also watched the video of the sessions and found no significant difference in number of disagreements between the two groups. However, when I asked participants to recall how many items they had agreement and disagreement on in each steps of their discussion, participants in the ACTIVE condition reported significantly higher number of items they agreed on, than participants in the PASSIVE condition. Participants in the PASSIVE condition also reported the number of items they had disagreement about significantly lower than the intervention group.

While there is no significant difference in the average perception of conflict and conflict management, between the two conditions, I noticed a trending difference in response to this single item: 'How frequently did people on your team disagree regarding the work being done?' in the scores for PASSIVE (M=2.21, SD=1) and ACTIVE (M=3, SD=1) conditions; $t(24)=-1.9$, $p = 0.06$. After observing the higher conflict recall in the ACTIVE condition, I watched the video recording of the meetings again looking for the frequency and quality of conflicts. I noticed the disagreements were relatively stronger in the ACTIVE condition and it can explain why people recall them higher than the participants in the control condition. I had no control on the difference in the number and level of disagreements between the two groups, yet it could affect overall experience of the participants. As described in the method section, I kept everything except the robot's disagreement management intervention, the same in both conditions, but I could not control the disagreements people initially have about items.

Regarding the efficiency of the meetings, overall I did not find any significant difference on the composite scale between conditions, however I found a (close to) significant difference in evaluating how wisely the group used the time for PASSIVE (M=6.8, SD=.36) and ACTIVE (M=6.1, SD=1.1) conditions; $t(24)=2.06$, $p = 0.050$.

7.5.1 Active Listening

To compare the disagreement management interventions, we looked at participants self-reports on some of their active listening behaviors and found significant and trending differences between the participants in PASSIVE and ACTIVE disagreement management condition. An independent t-test on single items showed significant and trending differences between the participants in PASSIVE and ACTIVE disagreement management conditions on their active listening skills and behaviors.

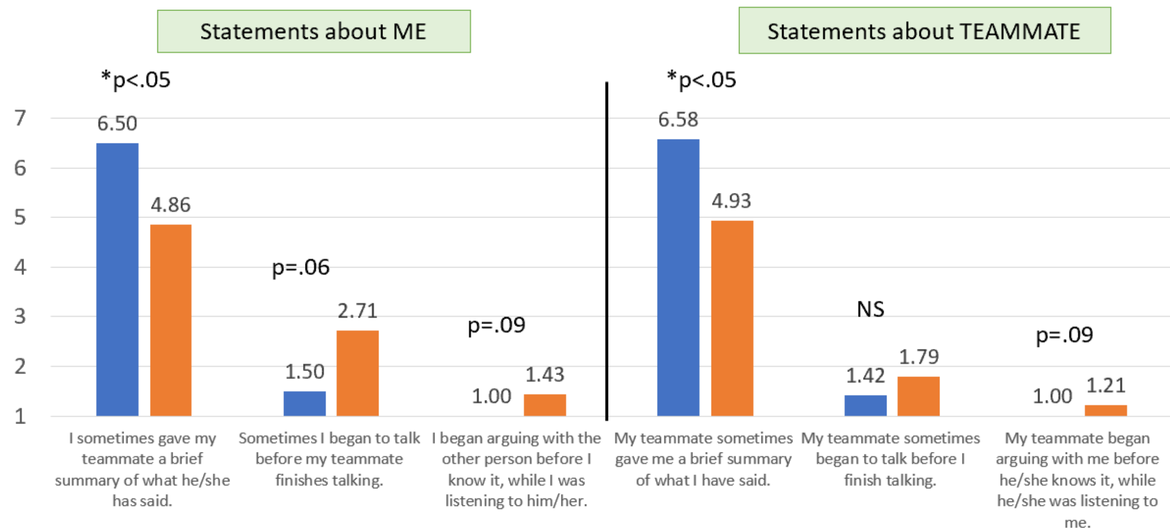


Figure 12. Participants rating of their and their partners' active listening behavior during the group discussion

7.5.2 Robot Evaluation

As described in section 7.4.3 participants answered several single item questions, as well as a composite scale on their perception of the robot's usability, personality and rapport. Among the single item questions, we found that participants rated the robot's friendliness and trust significantly lower in the ACTIVE condition. Table 11 shows the mean and standard deviation for each group.

Table 11. Robot Evaluation comparison between two conditions

	ACTIVE Mean (SD)	PASSIVE Mean(SD)	p-value
Rapport	5.49(1)	6.26(.6)	.024
Trust	5.33(.88)	6.5(.94)	.004
Friendliness	5.92(0.9)	6.71(.61)	.003

Last but not least, in my baseline measures I asked a few questions to assess individuals' feeling towards the group work (e.g., 'I enjoy working within a group.', 'I prefer to work within a group rather than work alone.') on a 5-point scale. I noticed that participants in the ACTIVE condition had significantly lower (t(24)=2.26, p = .03) responses to these two questions (Mean = 3.3 (SD = 1)) compared to participants in the PASSIVE condition (Mean = 4.1 (SD = .63)). This initial difference in feeling towards group work may explain some of the results I showed above.

7.5.3 Conclusion

I designed and developed a group facilitation robot that could provide interventions to manage disagreements and conflicts in a group meeting. 26 participants (forming 13 small groups) were recruited to attend group decision-making sessions facilitated by the robot. I conducted a between subject user study to compare two different disagreement management strategies delivered by a robot (ACTVE, vs. PASSIVE). My quantitative findings showed people were satisfied with their experience and liked the robot to be involved in handling their disagreements. They also appreciated and followed the active listening practices instructed by the robot. However, I did not see the ACTIVE disagreement management translate to much different in group performance and my quantitative results left me with a couple of unanswered questions such as:

- Why the robot with ACTIVE DM is rated as less friendly and trustworthy?
- Why participants felt the time was used less wisely in ACTIVE DM?

Furthermore, wanting to find out about the expectations from the robot and the applications of its facilitations in a real work setting, motivated me to conduct a focus group with coworkers in a real work setting and perform in-depth qualitative analysis to find out about their expectations from a virtual meeting facilitator.

Chapter 8: UNDERSTANDING THE DESIGN REQUIREMENTS OF SOCIAL ROBOTS IN WORKPLACES

To identify opportunities for improving the design of the group facilitation robot, I conducted semi-structured interviews with the participants in the disagreement management study (Chapter 7) and held a focus group with employees who are currently working together in an organization. I designed the interview questions and the focus group to answer my second research question regarding the expectations from and perception of facilitation robots at the workplace. I also wanted to find out more about the minor dissatisfactions reflected in the quantitative results of the user study (section 7.5). The qualitative findings of these focus groups demonstrated the feasibility and benefits of certain conflict management interventions. The study results also characterize a strong correlation between the unique characteristics of each group on the expectations from a facilitation robot. I discuss implications for design and how my findings challenge some of the features deployed in the current prototype.

8.1 Focus Group Methodology

To gather more insights into the group meeting experiences of information workers in the workplace, and to obtain a more realistic perspective of the desired capabilities of a group facilitation robot, I a series of interviews and focus groups in two rounds. The focus of this work was to explore how a conversational robot should be designed for group facilitation and how it can address the various challenges in a group decision-making setting I conducted 13 focus groups of 2 members with people who attended group decision making sessions facilitated by the robot (Chapter 7) and ran one focus group with 7 employees with strong group collaboration experience in the workplace, to understand their unique challenges and expectations for a conversational robot for group facilitation. To distinguish the two different types of focus group and for clarity, in the reset of this document I refer to the 13 focus groups of the user study as ‘user study/interviews’, and use the term ‘focus group’ to refer to the discussion with 7 employees in the workplace.

The focus group participants work in the same group at the administration department of a university. The average age of focus group participants was 42.8 (SD = 10.4), and four of them were male. The length of their experience is this team was between 0.5 and 7 years, and had regularly attended group meetings together. 50% of them said they regularly attend group decision-making meetings as a participant. In the focus group I asked members of a group at a workplace to provide input on their preferences for, and attitudes towards a conversational robot at the workplace. This process yielded crucial input from my target population, resulting in identification of key factors that impact people's interaction with robots in work groups.

I designed three activities in the focus group session to

1. collect participants' immediate feedback about the concept of using a robot to facilitate a group meeting,
2. learn about their *ideal* group facilitation robot and its desired functions to help with their common challenges in meeting (e.g, conflict), and
3. understand their perception of and reaction to the robot's failure and malfunctions during a meeting.

Northeastern University's IRB approved all activities, and participants were compensated for their time.

To start the session, participants first watched a 3-minute demo of the group facilitation robot while she introduces herself and describes some of her capabilities. They also watched a short video clip of a small decision-making session facilitated by the robot to see how it interacts with a group. After the introduction, they were engaged in a discussion about their overall impression of the robot, and what they think about the robot's role in a meeting.

In the second activity, I specifically asked participants to picture their ideal robot, who it represents, and what it does. They were instructed to fill out a questionnaire with six open questions about common challenges in their group meeting, the role of their ideal robot in a meeting, and how their ideal robot supports their meeting. This activity allowed participants to think about different services and capabilities of a facilitation robot in a group setting to resolve some of the group meeting challenges. The idea here is to provide room for participant feedback by not making them feel that what they had seen was a final product. After answering the questions, to widen how they think about the scope of the robot, I asked them to particularly discuss how the robot can help with the agenda deviation, dominance, and disagreement challenges, and to what extent the robot should be involved in the decision-making.

Next, I displayed video clips of four types of errors by the robot during a group meeting. The video clips were taken from my pilot study when the system development was not flawless yet, and the robot had some errors and failures. The robot's failure shown in the clips were 1) error in social interaction, 2) delay in responding and repeating, 3) not understanding what users say- make them repeat multiple times, and 4) misunderstanding what users say and move on with wrong consequences.(e.g. they say they want to keep a resume, she thinks they want to eliminate). Finally, I asked participants about their perception of errors in individual vs. group interactions, and how they think the robot should recover from a failure.

8.2 Findings

Inspired by grounded theory analysis [43], I conducted a thematic analysis on the 13 focus groups interviews of the user study described in chapter 7, as well as the focus group data with 7 employees,. Using the Atlas.ti 7 software, I coded transcripts, and labeled emergent themes and phenomena in the interview and focus group transcripts. In this section I report the main themes extracted from the user study interviews and the focus group discussion.

8.2.1 Feedback on the Current Facilitation Functions of the Group Facilitation Robot

I first discuss the participant's impression of the group facilitation robot, its facilitation functions, and how they believe such a robot can be helpful in mitigating some of the group meeting challenges. The participants found SoFi *'friendly'* and *'very interactive'* mainly '*because of the way the robot looked at [them] and nodded and said "uhum"'* during their conversation. They felt that SoFi was attentively listening to them and that made them feel they wanted to talk more. They also found the robot to very helpful as a meeting facilitator and mediator. For example P2 in Group 1 said:

> *"In general I was really impressed with the robot and it seemed like a really great facilitator of meetings. It did seem like the robot was mainly there to provide structure and background information on the items rather than inputting the robot's own ideas. So that was an interesting dynamic."*

Many participants also mentioned they felt the robot's presence like "*she's not a robot, [rather like] some person is sitting besides us and just having a discussion*", a third person in the meeting who added a *'sense of formality'* to their meeting. They also liked SoFi's human-like appearance and expressive facial expressions.

During the interviews and the focus group discussion, participants brought up a lot of their previous group meeting experiences to evaluate SoFi's capabilities. They remembered the common challenges they used to experience such as dominance, bias, topic deviation and not enough listening. The participant talked a lot about how some of SoFi's characteristics could be helpful in handling these challenges.

8.2.1.1 Balancing Participation by SoFi

All groups appreciated SoFi's role in balancing the participation and giving all members the chance to express their thoughts. They discussed how the robot encouraged them to share more thoughts and how it was helpful both in cases of agreement and disagreement to learn more about the members' ideas. Most groups brought up their previous experiences in group meetings where dominance and imbalanced participation was a major challenge and acknowledged how SoFi could be helpful in such situations. For example Group 13 said:

> *"I think this is a good program when you have a big group. It is always like a few people whose voice had always never heard and then like this other people who are more assertive and aggressive and then you know, people, the entire team goes for three are saying rather than being based on what facts are. So I think if you have like this kind of is a middle man who kind of lets everyone talk... This is the main difference that like you are, everyone can get a chance to talk, but in, in a normal meeting, like the people who are more assertive or more extraverted, they would let dominate the meeting"*

In the same time some groups mentioned that it is not always practical to expect equal participation from all members of the group. Especially when the group hierarchy is not flat and people with different expertise and seniority are present. They proposed that balancing participation can be always helpful, but the robot needs to consider peoples' roles and allocate sufficient time based on their role and expertise.

8.2.1.2 Meeting Structure and Time Management by SoFi

All groups found the session "*well-managed*" and "*goal-oriented*",and found SoFi very helpful in managing the structure and time of the meeting. They discussed how this enforced structure by the robot prevented topic deviation (which they said is a big challenge in their previous group meetings) and improved the meeting efficiency. Here group 8 discussed how the robot guides them through each step.

"I think what I liked the most was just how the robot facilitated our discussion. It was definitely very structured and we didn't go into a blind, which is very nice. she had an idea of how we were going to have this discussion and we followed it very well."

While all participants appreciated SoFi's role in enforcing a structure to the discussion, some groups affirmed they would have had more room for free conversation without SoFi, and that may result in a smoother conversation in some cases. Group 9 pointed out that if the robot forces a structure people may not feel free to discuss their own ideas and that may add to the tension when they have disagreements. They said:

"I think it [structure management] is helpful in order to keep the conversation calm and give like some time to actually listen to the other person and be a structure. But at the same time I may lack some dynamics in the conversation or cause sometimes it's easier to go back and forth on argument as long as there's respect and so for me it would, if it was something that we disagree more, I would ideally have her not in like not talking to every in between every time we talk so we can like actually like go back and forth and if there's a moaning, which we actually like not listen to each other, then she would take place. So you would prefer to have more time for free discussion and then at some points when you have a disagreement or something she can enforce the structure too."

8.2.1.3 Active Listening

Groups in the ACTIVE disagreement management condition felt that practicing active listening was the most helpful activity in resolving a disagreement and "*encouraging receptiveness*" that they wouldn't have done without SoFi. They affirmed that practicing active listening had reciprocal benefits of reducing the chance of misunderstanding the other person, and feeling validated and heard.

[P13]: "Those were good parameters that ensured that we listened actively because when you know that you have to recalculate and you have to go back and convey what you understood, like I said, it could clear any miscommunication. in the absence of that, the con of that is you end up not listening to that other point of view and do much stick to that."

[P1]: "Encouraging us to listen to each other, repeat back what the other person said, I feel like that made me feel more validated when I was saying, [because] like the other person was listening actively to what I was saying and that, we would be able to work together."

Several groups including Group 10 reviewed not enough listening as a prevalent problem in many group meetings at the workplace and mentioned how SoFi could mitigate this problem.

"P19: So whenever we go in a group meeting, when we have some discussion at work or at university, so everyone's just keep on talking so it doesn't feel like everyone can understand what they're talking about. There's no active listening and like I really like the

> *way she said, like whenever you, the other person is saying something, you should acknowledge it like by saying it again. So that was really different thing for me."That's very helpful because the other person feels like, yes, you are being understood. P20: Yeah. Yeah. So previously, like whenever we have some kind of group conflict. The other person is kind of very adamant about, what they're their own points. So they are like, even if you're talking, they'll be like, okay, like do you miss this thing about the point they want to say. So the way that robot made sure that we repeat the other person's thing wouldn't like it made us listen to the other person and the other person was also like okay. She got my point. So she's considering it."*

One important suggestion to improve this intervention was to adjust the frequency of the active listening practice. Groups who had multiple disagreements found it a little repetitive and time consuming to go over the whole process again and again.

8.2.1.4 Informational Support by SoFi

Participants in both the user study and the focus group envisions the information provision as an essential application of a group facilitation robot. The user study participants found the information that SoFi provided about the items 'very effective' in their decision-making process, and they specifically found this conversational way of information communication more engaging than receiving it printed out on paper. They also proposed that SoFi can play an important role in facilitating the decision-making process by providing extra information and facts to guide the discussion. For example P10 in Group 5 said:

> *[P13]:"I think SoFi gave us a lot of information, which we didn't think about. Like for instance, using the pocket mirror for signaling that is not something we thought about. Um, so if I had more information regarding the different options, I think I could make a more informed shots. So in that way, I think it as a great asset in terms of, you know, just giving us some background information which could then affect our decision making."*

While some groups said SoFi can provide this extra information as *"her input"*, four groups suggested that SoFi can present this information as other groups' or human experts' suggestions, *"instead of it having her chime in"*. They argued that the information and reasoning would sound more reliable and acceptable if the robot presents them as other people's opinion. Group 11 discussed suggest this as follows:

> *"maybe she could talk about previously some other group did this, some other information. If she doesn't want to give a biased thing, [just] some more information, through which we could solve the problem."*

Our findings show how the robot's level of information and knowledge could characterize its role in the meeting. When describing SoFi's role in the meeting Group16 found SoFi 'a bit domineering' and mentioned that SoFi's position in the group *'was a little elevated because she had so much knowledge'*. People would associate more intelligence and expertise to the robot if they feel the robot has a lot of information about the task.

Another concern regarding the information provided by the robot was the availability, accessibility, and the source of information. While participants of both the user study and the focus group found informational support as an essential application of the robot, a few groups mentioned that background information including *'industry standards and industry culture'* is not always available (neither exists nor is accessible by the robot) at the workplace. The robot would be helpful in situations that expert knowledge already exists, but in a realistic setting a robot *'might not have that institutional knowledge that the other people at the meeting would have.'* and it would not be useful in the role of an informational provider.

They also discussed the information type and the information amount they would expect the robot to supply, depending on the group members knowledge about the decision-making task. If group members know the domain well, the robot may have a bolder role in structure management, while if people are not very familiar with the decision-making context, the robot is expected to aid the group by providing information.

The participants commented on whether they would like the robot to be proactive in providing information, or it should wait for participants to ask for it. Most groups demonstrated a strong desire to be able to ask SoFi for more information about the decision task. The request to informational support has been also echoed in prior work in organizational sciences. Bercovitch described how a third-party group facilitator "becomes the information-gathering instrument and a "resource person"" [12]. Regarding the proactive information supply by the robot, considering that any information may not be useful for any group, they suggested the robot should confirm whether people are interested in the information before providing it.

8.2.2 Robot in Groups with Hierarchy

Another recurring theme was the hierarchy in group meetings and the robot's role in groups with hierarchy. The participants repeatedly described challenging experiences in *"meetings where there is kind of hierarchy and seniority",* where handling dominance and participation in these cases is not always straightforward. The power dynamics in a group with hierarchy may impact the decision-making process and outcome, in two ways: one issue is that usually meetings are *'dominated by the people with more seniority because they're the people leading the meetings'*. The second matter is the effects of group hierarchy on other members' feeling and contribution since *"normally people tend to agree with what senior people said and they don't speak up properly"*. Group 6 discussed their experience of such situation below:

> *"Some time it happens in a team meeting when we are like a group of four or something and they're like two senior managers sitting in the same meeting hall. They sometimes are just not able to come to a decision and we being the junior employees, we are not really able to like push our points much. Even if we feel that one of them is better because our decisions can always be considered as a biased decision because we are one of the two managers' team members. So at that time, if SoFi can speak, it will definitely help to resolve the issues."*

8.2.2.1. Balance participation based on members' roles

In such settings, *equal* participation is neither practical nor desirable, while interventions to encourage the involvement of all members and make every member's voice heard is highly essential. Participants affirmed the significant role that a robot can play in meeting with hierarchy as a *'neutral third party'* and insisted that the robot's role in the meeting, the language it uses, how it calls the members, and some of its functions should be modified in groups with a hierarchy to adjust to the dynamics of the group. Several groups proposed that instead of promoting *equal* participation, SoFi should be aware of people's roles, interest in the topic, and power status in the meeting and *'assign different levels of priorities to different people in the meeting.'* Group 1 said:

> *"In a meeting where there is a hierarchy, the robot can allot a certain amount of speaking time to the more senior people while still allowing time for the more junior people so that they're still able to express their opinions."*

Group 9 proposed that the robot can handle dominance in a group with a hierarchy in a respectful manner. The robot needs to be aware of the roles, powers, and relationships and recognizes who is dominating and ask the other members how they feel about the issue. Or recognize if someone completely neglected the significant points.

> *"If I like completely neglected some major point of yours, I think it would be helpful for it [SoFi] to not be like, why didn't you discuss this or like it up. But maybe like what do you think about this point as well? And maybe not even phrasing it as like this is what he said. Like you can acknowledge it but being like, what are your thoughts on this? Or like how do you feel about that? rather than like blatantly saying like, this person has a really good point. Why are you completely ignoring it? I think maybe just like asking a question that maybe he would have rather brought up if I'm his superior that like he felt like he couldn't challenge me a second time."*

8.2.2.2. The robots' role in meeting with hierarchy: Peer vs. Superior

Most participants reported that to them SoFi was more like a peer in their meeting, and they found it appropriate. One group described SoFi's role '*somewhere in the middle of a superior, and a peer, ... more in the lines of a mediator . calm and peaceful mediator*.'

When they were asked about the ideal robot's role and power status in groups with hierarchy though, one interesting observation was that multiple groups in the user study stated that they would prefer SoFi to represent a peer when the group hierarchy is flat, while they would want SoFi to be in a more superior role when there is a hierarchy in the meeting. Group 9 discussed this:

> *"P19: So if there is a feeling of like I'm superior, or he is superior than I, and we will have this feeling, then I'll make the robot even more superior. so that we are kind of like all equal in the sense of like we're all behind the robot. but if we're all equal I would just make her be like included in the group as another peer…. I think the way I'm going to feel more comfortable expressing my opinion and giving all my thoughts and saying, all I want to say is when I feel equal. So if I feel I have less power than you intend to saying what I think, I wouldn't say. And so the moment the robot, it's above and treats all of us at the same level. I think that would make me more relaxed. P18: And I feel that if there's someone superior than my boss and if I have some disagreements, like there's someone above who can listen and express and enforce some decisions on the boss that everything is fair. "*

Participants in the focus group discussed robot's role from the group managers' (or any member with high power status) perspective too. In contrast with Group 9' idea, one participant in the focus group mentioned that a boss or a manager might want the robot to take a more assistive role as opposed to a more superior role in the meeting. A group manager would think that he/she is able to moderate a meeting and enforce an agenda. So he/she would like the robot to help with administrative tasks such as taking notes, reminding times, or summarizing the meeting. Thus my participants suggested that in a flat hierarchy like a *'staff meeting'* where more decision-making is going on, the robot can represent a more authoritative role to enforce the structure and manage the time, while when a boss or a manager is present at the meeting, the robot can play the role of an assistant.

> *"I'm not sure this robot is best [to help in] in certain situations? Certain meetings, Mary [Department's dean] for one I wouldn't put this robot in her. I think it's a, if it's a, we need to come to a decision on something. Yes. But I think if it was a hierarchy of, you know, senior leadership, I don't even think besides recording and kind of taking notes, I don't know if the robot would be ideal. I think that it's just one of those types of situations that you just leave it be because they're all going to have that dominance and you know they are senior leadership, so you're not going to have them, you're not going to tell them to be*

quiet for one and you're not going to turn to the one person. That's just kind of taken it all in and analyzing it to be like, how do you feel? Because then you might not want to see it."

Taking a deeper look at the examples that focus group participants were discussing, I realized that when talking about meeting with hierarchy, they were specifically thinking of a high-level report meeting. A meeting where, for example, the department's dean is present, and less group discussion is happening. This better justifies the focus group reasoning for the robot's role in different meetings.

8.2.3 Level of Robot's involvement: Mediation vs. Arbitration

One critical aspect of having a social robot in a group meeting to assist the group is the extent to which the robot should be involved in various group tasks. A group facilitation robot can be a passive assistant to tell the time or query a data when a group member requests, or it can be actively involved in the decision-making task by using the outcome of analytical programs and have its own opinion in the group. Participants had different expectations of the level of robot's involvement in the decision-making activity. Some groups clearly wanted SoFi to have her own opinion and provide more analytical input in the group decision. For example, Group 5 said:

> *"The robot here should not be treated as a robot. I guess it should be treated as a person who has his own likings or dislike or whatever function it has. And then it's taking a decision of, obviously there is information the robot is going to have some informed decisions because it's not a human is going to get all the data and then it's going to take a decision. pros and cons or whatsoever it is. So in that case, I guess the robot should be treated as a person. Like this person has this opinion, this person has this and this person has that. So whatever goes good, that should be treated rather than feeling that it being biased because it's not going to be biased on our decision."*

In contrast, other groups strongly believed the robot should not have her own opinion and should be only a moderator and let the human be the *'final decision-maker'*. They *'preferred to at least believe you're making the choice'*. They also argued that people would assume that robot's input/opinion is based on analysis and facts and would take that for granted and agree with that. For example, Group 11 said:

> *"P22: If she'll get involved then people will tend to take his choices. So it's better that she is just guiding us and not getting involved in it. and it's good that she asked us question and we give the answer. rather than she she giving her own opinion. P23: yeah. I agree with that. You get not like a third party thing but not have her own opinions about what's been discussed."*

And a third group thought the robot should have a say and be able to provide an advice on demand, but group would only take it as "*as a third perspective.*". Group 6 said:

> *"So decision should be with us only. But when we [inaudible] she should not interrupt in between but probably when we asked Sarah about do you think about this and after listening to all of us she should give her advise decision. It would be definitely unbiased because he's a robot. She should have her say. Like we shouldn't consider her say as the decision but we should be the ones who take the decision but she should have her say."*

Some groups also related the desired level of robot's involvement to the information it provides and requested more involvement going beyond data querying and providing analytical interpretation and opinions based on that data and other people's opinion. The level of the robot's involvement and it's role depends on how much information people have about the task.

8.2.4 Meeting Efficiency with the Robot

Given the task-oriented nature of group decision-making meetings, their efficiency is especially important. My participants had mixed opinions about meeting efficiency with SoFi depending on the meeting goal, the level of their agreement on decisions, meeting type, etc. Overall, many groups found the structure enforced by SoFi helpful for meeting efficiency and thought without SoFi the discussion *"might have gone on longer"* on irrelative topics, and they *"would have wasted a lot of time"*. They reviewed their previous experiences in group meetings where they passed the meeting time and had to leave meeting with no decisions because the discussion had gone off topic, and thought SoFi could help in such situations by managing the time and keeping the meeting on track. However some groups raised concerns about meeting efficiency with SoFi in certain cases where the discussion can not fit into a firm structure. For example, an *'idea creation'* or *'design'* meeting where *"there's no advantage or disadvantage"* and a lot of back and forth is needed. Participants also felt the using the robot may decrease the efficiency of a meeting when there are extreme cases of emotional disagreements, or when there are very few disagreements. For example, participants in Group 8 said:

> *"So yeah, in that situation, I think it was less time efficient because I think we agreed on a lot of the items. Um, and so in that case it kind of could have been just like a one, two, three, like list them off, um, versus giving explanations for every single one. Um, and it, but I also think it comes down to personality. I think I write and I speak pretty concisely and I try not to be very long winded. And so if you have people that, you know that about your team, maybe the robots a little inefficient to use, but if there's certainly people in situations that need to be cut off and if that's the case then she would be helped."*

Similarly, Group 11 pointed out that the process may be unnecessary when people are in complete agreement:

> *"If there's a big conflict then yes, something like SoFi would be very helpful. but If there's like a very minor conflict I think, just discussing the human will be much faster. Depends, I mean on the like depends on the conflict."*

One group though mentioned that it is useful to listen to other people's perspective even when you agree, as they may see the issue from a different perspective.

8.2.5 Disagreement Management by the Robot

User study participants commented positively on how the robot *"went in the depth"* and *"walked them through a process"* to manage their disagreements efficiently during the group decision-making. For example, Group 8 discussed SoFi's role in disagreement management as below:

> *P16: "I think her just like asking us to identify objects to discuss was definitely something that helped. Without her, I think we just have to run down a list of just each item and now it's importance. When a disagreement was found, she definitely took the right precautions or the right protocols to just have us discuss. I think without her, um, disagreements like that would definitely be a lot more spontaneous, a lot more disorganized. Just having her moderate, it was definitely very, made everyone a lot, much more professional."*

Participants in the ACTIVE intervention condition liked actions that the robot took for disagreement management, such as practicing active listening (6.1.3) and reviewing the pros and cons of each item. They found it particularly helpful when SoFi asked them to think about and discuss the advantages and disadvantages of the items.

On the other hand, many participants who experienced PASSIVE disagreement management by the robot felt the lack of sufficient action by the robot to resolve their conflict. This kind of feedback indicates the need for appropriate interventions by the robot about disagreements. Group 2 in the PASSIVE condition said:

> *"In this situation she doesn't have anything after she has stated that there is a disagreement. I thought that it was just stating the obvious because I know there is a disagreement ..., and she just told us that basically you're not on the same page. I think it is important to state that you have not reached a decision. But you have to do something about it because it cannot be left unsaid if in a group meeting, if you just leave as certain topics onset or you don't reach a certain conclusion, it's not going to benefit anyone. Yeah. So yeah, it's very important to reach a stage in a group decision."*

Participants in the PASSIVE condition recounted certain aids that SoFi could offer after acknowledging a disagreement, such as reviewing the pros and cons and providing information and facts which I describe next.

They allude to reviewing the pros and cons and one of the necessary actions the robot should take. For example, P1 in Group 3 said: *"If there is a disagreement, I think the robot should tell the pros and cons of both of the items and should support a little from her side so that we can end up on choosing the right one."* They also proposed ways to improve the review of items' pros and cons by the robot. Group 7, for example, would have liked SoFi to provide more analytical reasoning based on the pros and cons.

> *"What it could have possibly done is to weigh them probably. Say if the two items being compared, the strengths and weaknesses of each of the two items and if the weaknesses outnumber the strengths, you clearly have a winner. So yeah probably taking on those lines compared to just offering plain opinions. It could have a countering different points here."*

Some groups in both conditions also expressed their desire to have a *"checklist"*, an aggregated metrics to evaluate items, or a visual comparison table where they could compare and *"see both of them* [pros and cons] *in front of"* them.

Our participants also expressed concern over the exit solutions. To prevent excessive discussion when two people are not converging on a disagreement, they suggest having a compromise option that allows the group to move on even when they disagree (disagree and commit). In the current design, SoFi asks the group to move on to the next step if they still disagree after three attempts of resolving a disagreement. These comments suggested that the robot should offer appropriate "move on" options even earlier and just makes sure that people don't have anything else to say by asking them whether they want to add anything else. Group 8 never reached out to the three attempts and explained the suggestion as below:

> *"I don't know if she had an option for us to reach a compromise because it seemed like it had to be one thing or the other. If we were able to, if there was an option to reach a compromise, I think that would also be a lot more how would be very beneficial to both parties involved. It's not ideal, but it's definitely something."*

8.2.4.1 A Holistic Approach to Disagreement Management:

Our interview data further characterized the impacts of managing the meeting structure and encouraging balanced participants on creating a group environment that is more resistant to negative consequences of group conflict. Managing disagreements among group members is not only about taking actions after a disagreement occurred. Perhaps much more can be done to prevent a disagreement from happening at all. Managing meeting structure, and balancing the participation play important roles in creating such meeting environments that are less prone to disagreements and the following tensions by clearing the discussion ground. Group 9 discusses this:

> *"[SoFi] Formalizes the argument more compared to any other just plain human interaction argument, formalize the arguments. Right. And, at the same time I think having it can be helpful to keep the argument calmer compared to when it's just a human because it allows you to express your ideas. You're allowed to voice your opinions at a certain time. …. So yeah, again, as I mentioned, there was a structured way to carry out the entire argument. ... By helping people hear each other out at the same time they are allowed to raise their opinions and uh, see if it leads to a conflict."*

Several participants particularly reported that *"when [they] did have disagreements, it was good to have that structure and also be allotted equal amounts of time to contribute our thinking."* They also found it helpful when SoFi gave them *"very clear and crisp"* instructions on the ways to solve the disagreement. These comments show that participants specifically connected the meeting structure and equal contribution time to help with their disagreements.

8.2.4.2 Providing Information to Manage Disagreement

Related to my findings on the implicit impact of the robot's other functions on disagreement management, participants specifically talked about the robot's role in providing information as an effective way to handle disagreements. Many participants expected the robot to provide *'new facts or information'*, or talk about *'new dimensions'* about the issue of conflict during disagreement management. For example group 7 said:

> *"[SoFi could] mention a few facts that could be compared with what the other person said. See for example, if she and I had a difference of opinion on something, and I mentioned something that she mentioned something else and the Robot mentioned something about that item, a few things. And we could relate one of our points to her [robot's] points. So yeah, that would make more sense."*

A few participants also said the robot could *'correct'* them when they are saying something incorrect in an argument. A suggestion that was echoed frequently by my participants was that the robot could bring up new information by referring to other human's experience or expert opinion instead of saying 'I think'. Overall many groups feel that they need to have more information to manage disagreement, and the robot should provide this information either as her input or other groups' experience.

8.2.4.3 Robot May Disagree but Should Never Embarrass

Interestingly I found this common pattern in the interviews that the robot may have to disagree with one side of the argument if it wants to manage the disagreement by expressing an opinion or offering information during decision-making. When we asked participants whether they would be offended if the robot disagrees with them, most of them strongly stated that they would not be offended, mainly because they do not associate any bias to the robot and consider the robot's opinion based on facts. Compared to a human disagreeing with them, they said they *'would be less upset and a robot'* and *'would think of it [robot's disagreement] less personally'* and would *'probably blame the [algorithm] if something goes wrong'*. Here is Group 4's response when I asked them whether they will be offended if the robot opposes their opinion:

> *"Not really because I know down the line that it's a robot. so definitely it's using its own as natural language processing skills that it's coming to some conclusion so definitely in fact I would appreciate it, because this robot is being trained with multiple teams and multiple things, so definitely I'm sure that the robots input is totally unbiased so there's nothing to feel offended about. I mean like for a person saying in something, you may feel that he has some interests with that guy. Maybe they are friends. That's why he's taking his side for the robot's decision would be based simply on the processing, which I can always like any other solution, ignore it if I don't want [inaudible]."*

One participant in group 8, however, felt that it is contradicting with the robot's neutral role if it disagrees with one side and she could be more offended at that case:

> *'I would be a bit more offended, If a robot took, personalized ticking. If a robot agrees with the person who I was disagreeing with, simply on the fact that the robot is supposed to act as a neutral party. with humans, that can be a bit more diff, different because humans are definitely open to being, having their opinions sway, but the robot is supposed to be objective and supposed to facilitate the discussion. If the robot took a side, It would be very unexpected.'*

Participants in the focus group also talked about the possibility of the robot bringing opposing information to one member's opinion. Some participants felt that *'they wouldn't want to hear that in front of the [group] meeting, and that would be nice to hear [it] afterwards'*. They felt although it may be less useful to have the information after the meeting, many people would not rather hear from a robot that they are wrong in front of their colleagues.

They also reflected on the likelihood of one developing a bias or a negative emotion, if the robot disagrees with him/her multiple times in a row. It would also '*create self-doubt'*, if people feel it is likely that the robot would disagree with them, and they would be discouraged to contribute to the discussion. They thought that it is important that the robot do not initiate disagreeing with a group member proactively. P6 in the focus group said:

> *"if I asked the robot if I was right or wrong and I was wrong, like that's fine. Like I'm asking it for its opinion or whatever, but if I'm sitting in a meeting right now, if it chimed in and said, you're, you're wrong, I'd be like the hell no, I'm not."*

I believe these comments confirm the people's sensitivity to their image in front of their group and the the importance of the robot's language and how it conveys the information that may contradict with one's opinion. It is critical for the robot to talk in a neutral way that avoids any personal misinterpretation and bias. As said by one of the focus group participants: "*the biggest thing you want to be careful of is embarrassing somebody in front of those other people"*.

8.2.4.4 Robot Has No Bias

Another prominent theme in my findings was participants' emphasis on the robot being unbiased. Most of the participants found SoFi *'really good at maintaining a neutral position'* and explicitly commented on how they see this as a valuable feature for a group facilitation robot.

Many participants communicated about how the robot's unbiased manner can help with different group challenges, such as dominance and imbalanced participation. For example, Group 8 discussed that robot could distribute the time among group members properly without *'bias of the role they play in the workplace'* and "*it just makes people that are more subordinate and more subordinate roles and more comfortable to say what they think."*.

Multiple participants also compared the robot facilitator with a human facilitator and found the robot more *'objective'*. They felt *"with the robot people would be more comfortable because not getting judged"*. Regarding robot's role in managing intragroup disagreements, they noted they could trust SoFi because they believed her information is unbiased and she is fair in her task. They also said they won't be offended if the robot disagrees with them because *'I'm sure that the robots input is totally unbiased so there's nothing to feel offended about'*. Given this feature of the robot, some participants suggested that it could be specifically useful for cases that could involve human's emotions such as 'exit conversations' where an employee of human resources should inform an employee that they can not work there anymore.

8.2.4.5 Robot as an Emotion Stabilizer

An interesting theme extracted from interviews was about how the robot can be an emotion stabilizer in the moments of conflict among group members. Several participants noted that when conflicts and disagreements in the group makes them feel unhappy with another member, they would prefer to be angry at the robot rather than a human colleague. Group 7 said:

> *“I think if we were hostile towards each other, I would rather be more angry at the device [robot] than at you. I feel like if I had to yell back at the device then it would kind of take off some pressure of you. That way you don't also become combative at me.”*

The same participant proposed that at the moments of disagreement, instead of two people keep arguing the robot can ask each person what they think and then, even if people feel negative they would blame the robot not their peer. P14 in Group 7 said the robot can say:

> *"Maybe [the robot can say] like what do you think about this point as well? And maybe not even phrasing it as like this is what he [the peer] said. Like you can acknowledge it but being like, what are your thoughts on this? Or like how do you feel about that? Um, rather than like blatantly saying like, this person has a really good point. Why are you completely ignoring it? ...Just kind of picking up on who's talking more who's not.* ***And if I do sense like,*** *this robot's pointing out a lot more of their points and my points at least like I'm like blaming the robot. ... I've been [mad] at the robot rather than directly at him for speaking up too much.”*

They also talked about the application of such a mediator role for the robot in meetings with hierarchy where people may not be comfortable complaining to their superiors but the robot can reflect their voice.

> *“Like if he [the other member] was my superior, there might be things that I would want to tell him that maybe the robot can pick up on. if like the robot realizes that like I'm only saying one or two sentences and he's saying a paragraph, I'll be like, do you have any more thoughts?”*

These comments reflect a phenomenon similar to the Displacement of Emotions in psychology [65] introduced by Freud. Displacement is known as a defense mechanism ‘in which people tend to transfer their negative feelings from the original source of the emotion to a less powerful person or object.’ A common example of displaced behaviors is when people feel angry but cannot direct that anger toward the original ‘source of their frustration, they transfer those feelings to someone or something else.’ My findings show that a similar pattern could happen in a meeting where the robot is present. Inspired by this mechanism I suggest that the robot could be intentionally designed in a way to play the role of the less threatening third object to which people can transfer their negative impulses.

8.2.6 Robot's Errors; The impact, Group tolerance, and Repair Strategies

Similar to any multimodal, computational system, a group facilitation system is not immune to errors and mistakes. The errors may be caused by a minor or major malfunction in any sub-modules of the system from the input device, to speech recognition, to response matching, to the text to speech process, to the output device. Yet, certain features of a group facilitation system made us hypothesize that the performance of such a system and people's perception of it could be more sensitive to potential errors. First, a group facilitation system needs to support a live meeting in real-time and be able to respond immediately to changes and requests in a dynamic environment in which errors may have a domino effect. Second, the group nature of the context in which these systems are used may provoke different perceptions of the robot's errors compared to individual human-computer interaction.

Given the dynamics and challenges inherent in a group decision-making meeting, I sought to understand how participants feel about the robot's error in the interviews.

A couple of user study groups reported some errors in the interaction mainly pointing to the robot's listening and understanding capabilities, repeating and interruptions. One group for example said: *"it seems that there is a little discrepancy with what we were saying and what it was feeding back to us."*. As a repair strategy to limit the consequences of such errors, several groups suggested providing *"confirmation and summary"* by the robot. The study participants also reflected on the importance of having a transparent and clear decision-making process for the robot. For example, if the group facilitation program is eliminating an option, it should make sure to clearly describe what it is going to do and the reason behind it.

Our participants deeply discussed the group tolerance towards the robot mistakes depending on the meeting type, and the error type. For example, about the errors in different types of meetings group 1 said:

> *"Well, I think it would be different if it was an internal meeting with just people in our company versus if it was a client meeting. Like if we were using these with our external clients and then it got the client's name wrong. Um, I think that would be more frowned upon than if it was just a casual internal meeting where we were trying to just decide on something"*

Similarly, the focus group participants compared group tolerance towards the errors in a serious and urgent group decision making with stakeholders in a company to a less formal meeting with peers who may have more patience for a new technology.

Regarding different types of errors, participants felt 'voice recognition' mistakes and 'robot's delays' (either in understanding or responding) are the most frustrating errors that could cause *'missing points and getting off track'*, while they found interruption less annoying in a group decision-making sessions. They also talked about errors with practical consequences, for example an error in disagreement detection, or an error in detecting the option under discussion in a live conversation, may result in non-reversible wrong tasks by the robot with a high cost for the group. Several participants pointed out the necessity of having a repair strategy. As a potential workaround, one group suggested the robot to support commands like 'go back', skip, or 'rewind'.

Comparing the errors and robot's malfunctions in a group interaction versus an individual interaction with a robot, participants found it is a *'obviously that's bigger issue'* in one on one, because in a group '*the natural instinct is to just turn it off and just deal with the group'*. In the same time, they pointed to the importance of the individuals' collective self-esteem [122] in front of the group and the high cost of impairing it.

8.2.6.1 Building Trust over Time

Establishing trust in an automated robot is essential for effective human-robot collaboration[40]. Trust in a robot is particularly important when its tasks do not have immediate visible traces, and a human user should rely on trust to continue an interaction. When talking about using the robot in real group meetings, the focus group participants explained how they would develop trust over time to any technology including this group facilitation robot. For example, P6 mentioned the common failures people experience in using even a simple voice recorder. He then explained how people would need to inspect a new technology-driven device like a robot, and have trial and error to build a trust over time:

> *"It would take a while for me to feel comfortable with like asking the robot to take notes for me without somebody else. Also taking notes or myself take notes to kind of compare to see if it's like, if it's helpful for me, if it's the same kind of way that I would do it.."*

Overall participants' comments about the robot's errors in the group facilitator role and their tolerance towards it, show that the groups sensitivity to such errors may vary depending on the meeting type, meeting phase, and error type. In certain cases, the time and energy cost of repairing a mistake may be more than the cost of leaving it unrecovered. The group facilitation system should have a program to prioritize error handling methods in different scenarios to provide a flexible and adjustable repair strategy. For example, P5 in the focus group described a scenario in an important company meeting where *'some people speak very softly'* and the robot cannot hear them, and asks them to repeat or talk louder. P5 added:

> *" It's like that kind of disrupts the meeting more than it's worth because other people might be able to hear it or, or she could say, do you mind speaking up? And that kind of helps all the peppers. So it's all about the, in my opinion of how they do it. Not embarrassing somebody but trying to be constructive."*

8.3 Discussion

My primary objective in this chapter was to investigate the right design for and issues around a conversational interface -- such as a social robot -- for group meeting facilitation. Building upon prior work in organization management sciences I proposed an initial prototype for a group facilitation robot that provides interventions to manage the meeting structure and time, balance participation, and manage intragroup conflicts. I conducted interviews with user study participants and a focus group to explore people's perception of and expectation from such technology in their group meeting.

My findings characterize the ways in which a social robot can benefit a group meeting, and how it can help with dominance and disagreements in groups. The findings particularly shed a light on the role that a social robot could play, and the complex aspects of human-robot relationship in a group decision-making meeting.

8.3.1 Robot's Role in Group Meetings - One Role Does Not Fit All

An important research question I sought to answer in this study was to find out what role the robot should represent in a work group and what people would expect from it. My findings provide a clear answer; there is no single role fitting all groups. Although my participants saw great value in having a facilitation robot in group meetings, they did not converge on one best task for the robot. There is not a single ideal role or task for the robot as the group facilitator. Depending on the meeting type, the group hierarchy, the task, the time, and the group size, each group may expect the robot to provide certain types of facilitation.

The same is true about disagreement management. Previous research has suggested that handling a disagreement is related to the meeting type, the relational background among participants, the goal of the meeting, the hierarchy of the meeting, etc. [5]. The specific context and power dynamics of a meeting affect if not determine the way in which opposing views are expressed and should be managed. Thus, different groups may prefer to set particular rules and standards for the disagreement management interventions delivered by the robot.

Moreover, the robot's role in the meeting very much depends to the knowledge and information it has about the meeting task. If the robot cannot provide any further information, or solution, expectations from it may be limited to documentation assistance. However, by providing information, it will be seen as an expert (or superior role) that people look up to and assign more intelligence to it.

I suggest designing a robot with programmable roles and services, which any team can adjust based on its needs. Enabling groups to pick their desired features from a defined range of options would create more flexibility in using the robot, yet does enforce a structure to meetings. For example, even in a single group, people may have different ideas about the time allocation or the robot's level of involvement, but selecting a facilitation plan (from multiple options) would impose some regulations, without feeling mandatory.

8.3.2 Human - Robot Relationship

Individuals build a relationship with a robot and learn from its behaviors overtime. Consequently, robots' malfunctions in any interaction could create a bad reputation, and its smooth performance may create higher expectations [173]. In the group facilitation context, it is critical that the robot maintain its neutrality. A negative implication of forming a relationship history between a human and a robot is that it can cause a human to associate a bias to the robot after a single interaction in which the robot makes a mistake or disagree with him/her. For example, if the robot does not recognize one's voice twice in a meeting or ignore one's idea by mistake, that person may associate intentions to robot's behaviors over time or even self-doubt her/his abilities in a meeting. Individuals need to feel safe and comfortable to express their thoughts in a group. Thus, maintaining a positive human-robot relationship in a group is extremely important because a negative feeling towards the robot would impair one's teamwork skills and feelings towards the group.

I suggest that social facilitation of the robot can downplay the ripple effect of robot's malfunctions over time. Greeting and short social chat are good ways to create a positive vibe in the meeting. Providing personalized touches to each individual (e.g., calling them by names) can also be helpful in maintaining a warm and neutral image in the group.

8.4 Conclusion

I designed and developed a group facilitation robot that could provide interventions to manage disagreements and conflicts in a group meeting. I conducted semi-structured interviews with 13 dyads following their interaction with the robot, as well as a focus group with 7 participants, to evaluate the robot's performance and identify opportunities for improvement. The findings shed light on the right design and desired functions of a group facilitation robot in the workplace. Results also characterize different roles a conversational robot can take in a group meeting depending on the type of the meeting, the group dynamic and the group hierarchy. I discuss implications for design of a group facilitation robot, and how my findings challenge existing social robots at the workplace.

Chapter 9: CONCLUSION

Conducting productive group decision-making meetings while maintaining members' satisfaction is not easy. Group meetings are a prevalent form of gathering in the workplace to make decisions collaboratively. In fact estimates suggest that, on average, about a third of an employees' time is spent in these meetings. However, many of these meetings are not productive, resulting in wasting a considerable amount of time and money in organizations. Research in Management Science has identified the leading causes affecting the meeting efficiency as: deviating from the agenda, going off-topic, dominance of individuals (imbalance in participation), and intragroup conflict.

The same body of research has demonstrated that placing an individual in the role of a group facilitator can significantly improve meeting performance by alleviating the effects of the challenges mentioned above. However, the high cost of hiring a human facilitator, besides other logistic barriers, prevent many groups from employing one.

In my dissertation, I proposed an alternative approach for group meeting facilitation, and I hypothesize that it could mitigate the effects of many of these challenges. The core idea of this dissertation is to employ a conversational system (e.g., a virtual agent or a robot) to facilitate a group decision-making meeting by simulating some of the tasks of a human facilitator. It is important to note that this system never seeks to replace a human facilitator but to replicate some of his/her tasks when a human facilitator is not available—which is very likely as suggested by estimates. Building upon previous work that used robots and agents as peers in multiparty interactions, I developed a Group Facilitation Robot that can wisely guide a group to have a productive and efficient discussion by providing real-time interventions. To alleviate the common challenges in group meetings, the group facilitation robot seeks to manage the structure of a meeting, help manage the time, balance the members' participation in group discussions, and manage disagreements among group members.

9.1 Contribution

In this dissertation, I developed a framework that enables a conversational system to take the role of a group facilitator in a face-to-face group decision-making session (described in Chapter 3). Through an incremental development process, I developed three prototypes of group facilitation conversational agents, and conducted a series of user studies and focus groups to evaluate the effectiveness of such agents in improving people's group meeting experiences and managing multiparty interactions.

Inspired by prior work in group facilitation and group decision support systems, I first presented a framework for an automated group facilitation system. This initial framework declared the different modules a group facilitation system would need and the communication that would be required between these modules. In the framework, I also defined specific types of facilitation that the system could provide as well as when to offer them in the group meeting life cycle.

In order to assess the acceptance of a CA in a decision-making meeting and explore how it should be embodied, in my first study, I addressed a fundamental question: "*Does a conversational agent need a face to interact with a group?"*. I designed a Wizard of Oz experiment to evaluate the initial framework and explore the right embodiment for the group facilitation system. Results of that study highlighted the benefits of a human-like embodiment for the group facilitation agent to improve the perceived trust, rapport, and power of the agent. Qualitative findings also confirmed the feasibility of using a conversational agent as a group facilitator. Participants greatly appreciated the interventions provided by the agent and found them helpful in their group decision-making session. Furthermore, they requested more informational support from the group facilitator and commented on the importance of the unbiased nature of the agent. They found the unbiased and warm behavior of the agent could be an 'affective catalyst' to improve the social dynamics in a group. Findings of this study address my RQ0 and RQ2 regarding the feasibility and the suitable embodiment for a virtual group facilitator.

Next, I developed a fully automated group facilitation system that used multimodal sensor inputs (e.g., user gaze, speech, prosody, and proxemics), as well as inputs from a tablet application to intelligently guide a group decision-making meeting by providing real-time instructions. Results of a between-subject study of 20 user groups showed that the automated robot facilitator was accepted by group members and effective in enforcing meeting structure and balancing group participation (RQ3). My results also demonstrated that participants followed the robot's instructions during the meeting (RQ4).

I then decided to further focus on specific challenges that can arise in a group meeting and study how a group facilitation robot could handle those. I turned my attention to a common problem in meetings: conflict. I first sought to determine if conflicts could be automatically identified during a group discussion. I developed a machine learning model to detect disagreements in group conversations based on the verbal communication among members (language features of the utterances). I then extended the meeting facilitation robot to provide specific *decision-making* facilitation including conflict management interventions and informational support in a group decision-making meeting. Inspired by the group behavioral sciences, I designed a series of interventions to manage disagreements among group members, and I developed a prototype of the group facilitation robot that provide disagreement management interventions in addition to task-related information, and general meeting facilitation. I evaluated this prototype in a between-subject study. The quantitative results indicated that meeting participants followed the robot's advice when conflicts arose. However, there was insufficient evidence to indicate that these interventions were effective at increasing satisfaction. The qualitative findings were illuminating and should be beneficial to inform future designs of such systems. Participants greatly appreciated the disagreement management interventions by the robot and found specific features of the robot, such as its unbiased character, benefit its role as a group facilitator (RQ5). The qualitative analysis also shed light on what role the robot should play in different types of group meetings, what people expect from it, and to what extent it should be involved in the decision-making process. My findings also highlight the complex dynamics amongst members in groups with a hierarchy. The findings also indicated that participants had differences in their desired disagreement management processes and ideas regarding how a robot could be effective in addressing them.

This work contributes insights into the design of conversational robots and agents for group decision-making facilitation, informs efforts to support design practices in this space, and identifies opportunities for future research on social robots in group settings at the workplace. I also provide a design framework and discuss how it can be used for creating user-centered facilitation robots for group facilitation.

9.2 Design Guidelines and Lessons Learned

Here I describe overall lessons learned from developing and evaluating three group facilitation agents. I use the lessons learned from conducting three user studies alongside with many interviews and focus groups, to describe considerations that must be made when designing a social robot to interact with information workers at the workplace and facilitate group meetings. I also discuss what existing features in the prototype show promise within this context, which need to be rethought to meet the needs of different groups, and open questions that require exploration through further empirical research.

9.2.1 Make the Facilitator Tailorable to the Situation

Groups are as diverse as individuals. Each group has its own needs and dynamics based on specific features such as group tasks, meeting type, group size, and group hierarchy, and the responsibility and personality of its members. Even for a given group, the requirements and dynamics may vary from time to time. Consequently, different group meetings require different facilitators. Thus, designing one facilitation robot that works optimally for all groups is impossible. A group facilitation robot may need to represent different roles and provide different features to different groups. It needs to be designed with flexible options that each group can tune based on their needs. The participants clearly conveyed that they would want *'One type for each type of meeting'* when asked about their ideal robot's role. Although group meetings may have countless differences, various types of group styles can be categorized into fewer classes based on the same features that cause the variation (e.g., group task, size, hierarchy, formality, etc.). Therefore, designers can define a limited number of adjustable options for each feature (e.g., level of involvement, proactive vs. passive, equal vs. dynamic time allocation). I thus suggest that a group facilitation system should be designed with programmable roles and features and enable groups to choose the group facilitation type and level based on their meeting. Indeed, research is needed to examine what types of features are more and less effective at covering different groups' demands.

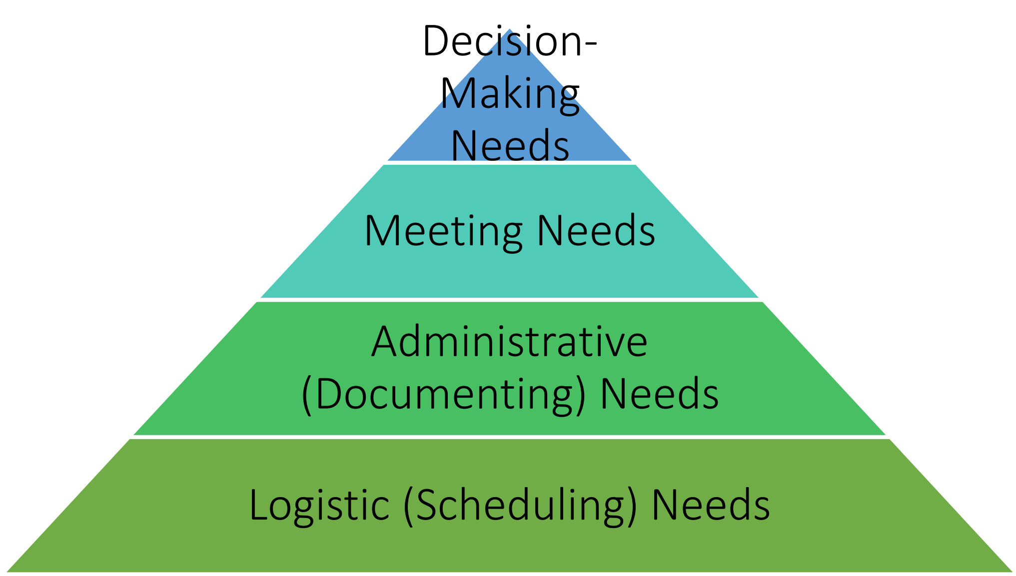
9.2.2 Provide More Informational Support

Most of participants commented on the importance of access to information during their decision-making tasks. Many of them were looking for informational support and the robot's *'ability to search for information'*. When talking about experiences of group decision-making, they recalled moments during a meeting when they needed a piece of information to act on, either to resolve a disagreement or to make a decision. They envisioned a significant potential in a robot that facilitates their access to information. They would expect a robot to be able to query information about the group meeting task and respond to their questions when they request more information. Comprehensive informational support by the robot requires it to have access to knowledge about the meeting task. A group facilitation robot also needs to know the discussion context to limit its online search space when people ask questions during a meeting, and it does not have the answer in its local core knowledge. Adding natural language understanding capabilities could result in great improvement for on-demand and real-time informational support by a facilitation robot.

9.2.3 Provide Logistic and Administrative Facilitation

Any group meeting requires logistical arrangements [56]. Someone needs to reserve a room, check the equipment, order refreshments, prepare notes, find available times and schedule the meeting, send calendar invitations, and so on. Surprisingly, these seemingly simple tasks can take a lot of time and energy. Our focus group participants, who all were experienced employees, expressed a strong desire to have a robot that can provide this logistic facilitation. Going beyond logistical assistance, they also requested administrative and documentation assistance. They would like the robot to help the group in tasks such as taking notes, summarizing the meeting, preparing a report, and identifying the next steps. They saw great value in a robot that can support groups in logistic and administrative preparation, as they believed planning and arrangements before and after a meeting could be as important and effective as the facilitation during a meeting. Hearing their needs lead me to picture the Maslow's Hierarchy of Needs for Group Meetings (Figure 14) with logistic and administrative as the basic needs of the group. The literature on group decision support systems (GDSS) reflects a similar taxonomy, by identifying information services, documentation services, and mediation services as three classes of required services in a GDSS [105].

Figure 13. Maslow's Hierarchy of Needs for Group Meetings



Our current prototype does not explicitly provide any logistic or administrative facilitation. However, given state-of-the-art technology and the existing integrated enterprise packages for IT services, incorporating these types of facilitation into a meeting facilitation system is quite feasible. More empirical research is needed to systematically explore more comprehensive group facilitation systems while supporting the various needs of a group.

9.2.4 Visual Aid, Clarity and Structure: A Holistic Approach for Disagreement Management

Often, disagreements can be rooted in factors that are not apparent in the first place. As identified by Bercovitch, communicational (conflicts arising from misunderstandings, etc.) and structural (conflicts related to organizational roles) factors are two of the factors that serve as sources of conflict besides personal differences of people [12]. Considering the impact of communication and structural factors, a group support system should aim to eliminate the effects of the root factors as much as possible. This requires providing a clear meeting climate, a structured meeting process, and an easy to use medium to access and share information.

As I showed in section 7.6.4, different functions that my current group facilitation system provides, such as promoting active listening, encouraging balanced participation, and enforcing a meeting structure, already help groups dealing with disagreements. Yet, managing disagreements by the robot can be further improved by focusing on informational support and providing a clear visualization of opinions during a discussion. In line with findings of previous GDSS research regarding providing features to remove communication barriers [53], my participants requested the information, and pros and cons of options to be displayed on a shared screen. Adding such features besides the specific disagreement management interventions can further enhance the system performance in handling conflicts in a meeting.

9.2.5 A Framework for Social Robots as Meeting Facilitators

Thus far, much of my discussion has centered on ways to improve a group facilitation system based on groups’ needs. Continuing this thread of discussion, here I propose a series of design guidelines on how a group facilitation system can be designed to provide supports for groups whilst not introducing additional burdens given the already challenging group dynamics.

Findings of my studies, echoed the discussion in previous work on Group Decision Support System (GDSS) regarding the three levels of GDSSs. As reviewed in chapter 2, DeSanctics and Gallupe [53] proposed a taxonomy of three levels of GDSSs. Level 1 GDSSs provide technical features aimed at removing communication barriers, such as large displays for sharing ideas and anonymous input. Level 2 GDSSs provide decision modeling techniques aimed at reducing uncertainty in group decision processes, often with automated planning and analytical tools. Level 3 GDSSs are characterized by machine- induced communication patterns with formalized procedural rules (e.g., parliamentary procedure).

Qualitative findings of my last study indicated the importance of Level 1 GDSS features such as facilitating scheduling a meeting, and sharing information during a meeting, beside Level 2 features.

Figure 15 shows a summary of lessons learned and guidelines for future research on group facilitation robots.

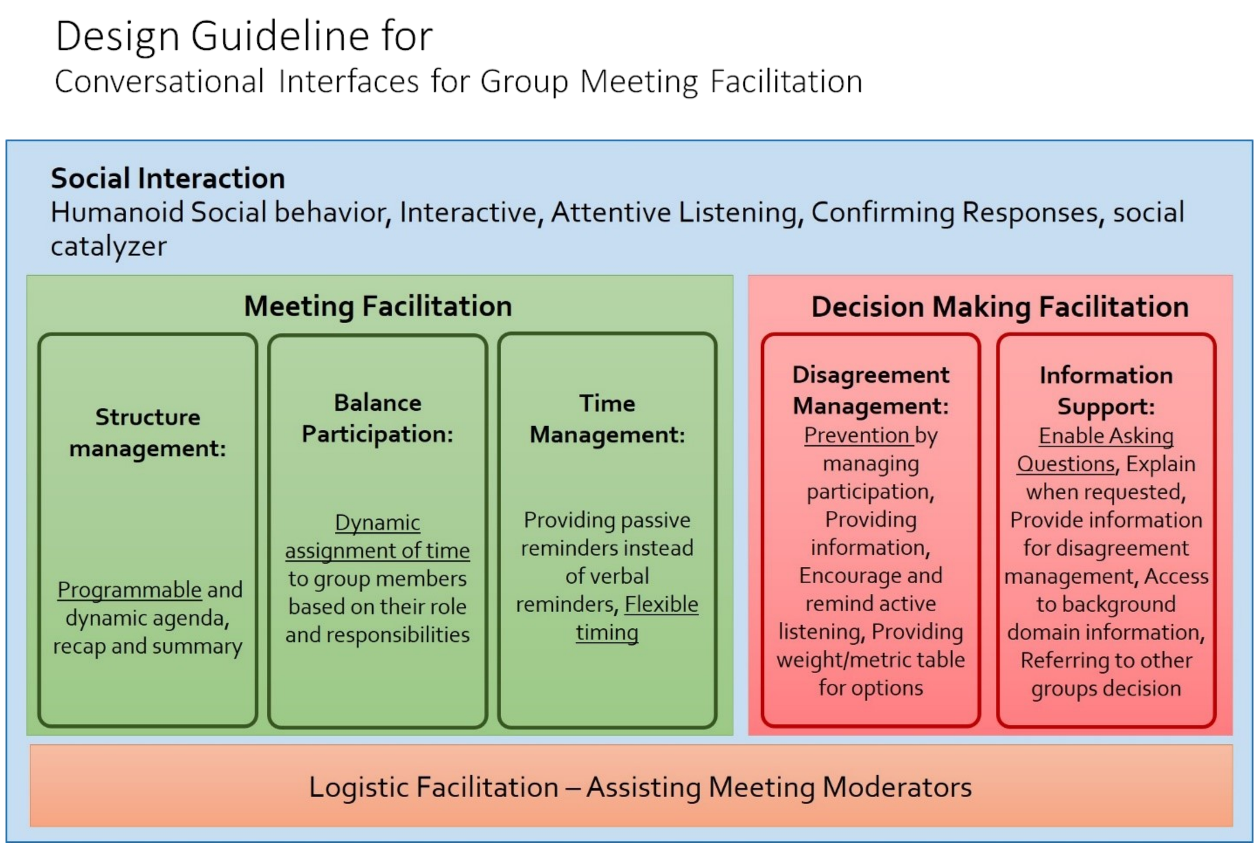


Figure 15. On overview of lessons learnt and my proposed design guidelines based on findings of my studies

9.3 Limitations

The experiments described in my dissertation have several important limitations beyond the small convenience samples of study subjects who participated in the studies.

In order to isolate factors that I wanted to study experimentally, the settings and tasks were very controlled, and thus may lack ecological validity. My task focused on collaborative decision-making and is similar to real-world problems in a variety of application domains including selecting conference papers, making funding decisions for research or investment portfolios, and accepting students to academic programs. However, unlike many real-world scenarios my task was short enough to be completed within a single session without pre-work done individually before the meeting.

Another major limitation of my experimental setting is the group size. To simplify the design process of the robot as an initial prototype I only studied groups with 2 participants which is the minimum possible group size. However, in real-world workplaces, on average between 7-15 people attend in most group meetings [164]. Moreover, in a real-world setting group members typically know each other and have a collaboration history in most cases, while in my studies I only recruited participants who did not know each other to minimize error variance.

9.4 Future Work

Although estimates by a survey suggest that most oral communications are dyadic at the workplace [152], other studies have found that the average meeting size is between 7 to 9 people [164]. An immediate next step to expand my research is to modify the developments and adjust interventions to be suitable for groups with more people. The rules and procedures in my framework are defined in a way to support larger groups. However, more empirical research is needed to test the system performance in larger groups, and examine how group dynamics and interactions in larger groups would respond to a group facilitation robot. I foresee significant modifications would be necessary to adjust the intervention timing and the robot's contributions to maximize meeting efficiency.

There is no doubt that a robot that understands the discussion and knows the context is more capable of providing relevant and effective interventions. In my system, I used basic approaches to natural language understanding (NLU), such as keyword detection, to make the robot understand the context. However, much more could be done to enable the robot to understand and grasp the intentions in a group conversation. Identifying the decision option that is being discussed at each moment, the final decision about each item, and the topic of a disagreement statement, are a few of many important functions that can be further improved using more sophisticated NLU techniques.

Given the prevalence of online group meetings in the post-COVID workplace, I envision that providing automated group facilitation services in online meetings holds great promise for enhancing group performance and satisfaction. Even though the physical presence of a robot will be missed in an online setting, the access to sensor data such as gaze and voice is much easier. In addition, online meetings can be leveraged to facilitate, enhance and expedite capturing meeting data (the first layer of my group facilitation framework [Figure 2]), which could result in more accurate and appropriate interventions.

Regarding automated conflict management, my work was a first step towards demonstrating the feasibility of using a robotic system to manage conflict in a formal setting such workplace meetings. Yet many steps could be taken to improve the performance and efficiency of these interventions. Looking at how human facilitators approached conflicts leads us to see the next steps to improve conflict management interventions by a robot. A potential next step is to classify interventions based on the level of disagreement. Another suggestion is to first identify the desired level of assertiveness and cooperativeness of each participants and then provide interventions to one of the five conflict modes introduced in Thomas and Kilmann conflict handling model (avoiding, accommodating, compromising, competing, and collaborating) [182].

Since expressed emotions are good indicators of argumentation status [97, 174], detecting emotions and involving them in the disagreement detection module could greatly improve conflict management interventions. A robot that is aware of meeting participants' emotional states and can effectively integrate this knowledge into its intervention could adjust its tone, attention, and non-verbal behavior resulting in a higher degree of user satisfaction.

Last but not least, in dispute resolution, a recommended approach by dispute managers is to communicate with each party separately [72]. The strength of this approach lies in the fact that it helps an individual to concentrate on his/her situation, needs, and alternatives without the fear of being judged in front of a group. A third-party consultant "remains detached from an individual, but his intervention, listening, probing, interviewing and explicit confrontation of the conflict issues, sets the basis for self-diagnosis and improved performance." [12]. A computer-supported group facilitation system could manage private communication with each individual via electronic messaging, and let them see the unnoticed aspects of an alternative in an unbiased manner. Research is certainly needed to examine the feasibility of this approach.

References

[1] 6 Steps to Resolving a Level 1 Disagreement: 2014. *https://managementhelp.org/blogs/facilitation/2014/07/15/6-steps-to-resolving-a-level-1-disagreement/*. Accessed: 2019-06-05.

[2] 55 Million: A Fresh Look at the Number, Effectiveness, and Cost of Meetings in the U.S.: *https://blog.lucidmeetings.com/blog/fresh-look-number-effectiveness-cost-meetings-in-us*. Accessed: 2019-05-13.

[3] Al Moubayed, S., Beskow, J., Skantze, G. and Granström, B. 2012. Furhat: A Back-Projected Human-Like Robot Head for Multiparty Human-Machine Interaction. *Cognitive Behavioural Systems* (2012), 114–130.

[4] Amason, A.C. 1996. Distinguishing the Effects of Functional and Dysfunctional Conflict on Strategic Decision Making: Resolving a Paradox for Top Management Teams. *The Academy of Management Journal*. 39, 1 (1996), 123–148. DOI:https://doi.org/10.2307/256633.

[5] Angouri, J. 2012. Managing disagreement in problem solving meeting talk. *Journal of Pragmatics*. 44, 12 (2012), 1565–1579. DOI:https://doi.org/10.1016/j.pragma.2012.06.010.

[6] Anson, R., Bostrom, R. and Wynne, B. 1995. An Experiment Assessing Group Support System and Facilitator Effects on Meeting Outcomes. *Management Science*. 41, 2 (Feb. 1995), 189–208. DOI:https://doi.org/10.1287/mnsc.41.2.189.

[7] Arora, S., Liang, Y. and Ma, T. 2016. A Simple but Tough-to-Beat Baseline for Sentence Embeddings. (Nov. 2016).

[8] Atkinson, R.K. 2002. Optimizing learning from examples using animated pedagogical agents. (2002).

[9] Bales, R.F. 1949. *Interaction process analysis; a method for the study of small groups.* Cambridge, Mass., Addison-Wesley Press, 1950 [i.e. 1949].

[10] Baron, R.S. 2005. So Right It's Wrong: Groupthink and the Ubiquitous Nature of Polarized Group Decision Making. *Advances in experimental social psychology, Vol. 37*. Elsevier Academic Press. 219–253.

[11] Baylor, A.L. 2009. Promoting motivation with virtual agents and avatars: role of visual presence and appearance. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 364, 1535 (Dec. 2009), 3559–3565. DOI:https://doi.org/10.1098/rstb.2009.0148.

[12] Bercovitch, J. Conflict and Conflict Management in Organizations: A Framework for Analysis.

[13] Bevacqua, E., Pammi, S., Hyniewska, S.J., Schröder, M. and Pelachaud, C. 2010. Multimodal Backchannels for Embodied Conversational Agents. *Intelligent Virtual Agents* (2010), 194–200.

[14] Bhattacharya, I., Foley, M., Zhang, N., Zhang, T., Ku, C., Mine, C., Ji, H., Riedl, C., Welles, B.F. and Radke, R.J. 2018. A Multimodal-Sensor-Enabled Room for Unobtrusive Group Meeting Analysis. *Proceedings of the 20th ACM International Conference on Multimodal Interaction* (New York, NY, USA, 2018), 347–355.

[15] Bickmore, T. and Cassell, J. 2001. Relational agents: a model and implementation of building user trust. *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '01* (Seattle, Washington, United States, 2001), 396–403.

[16] Bickmore, T. and Cassell, J. 2005. Social Dialongue with Embodied Conversational Agents. *Advances in Natural Multimodal Dialogue Systems*. J.C.J. van Kuppevelt, L. Dybkjær, and N.O. Bernsen, eds. Springer Netherlands. 23–54.

[17] Bickmore, T.W., Pfeifer, L.M. and Jack, B.W. 2009. Taking the time to care: empowering low health literacy hospital patients with virtual nurse agents. (2009), 1265.

[18] Bohus, D. and Horvitz, E. 2010. Facilitating Multiparty Dialog with Gaze, Gesture, and Speech. *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction* (New York, NY, USA, 2010), 5:1–5:8.

[19] Bohus, D. and Horvitz, E. 2011. Multiparty Turn Taking in Situated Dialog: Study, Lessons, and Directions. *Proceedings of the SIGDIAL 2011 Conference* (Stroudsburg, PA, USA, 2011), 98–109.

[20] Bohus, D. and Horvitz, E. 2011. Multiparty Turn Taking in Situated Dialog: Study, Lessons, and Directions. *Proceedings of the SIGDIAL 2011 Conference* (Stroudsburg, PA, USA, 2011), 98–109.

[21] Bostrom, R., Anson, R. and Clawson, V. 1993. Group Facilitation and Group Support Systems. (1993).

[22] Bousmalis, K., Mehu, M. and Pantic, M. 2009. Spotting agreement and disagreement: A survey of nonverbal audiovisual cues and tools. *Proceedings - 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, ACII 2009*. (2009), 1–9. DOI:https://doi.org/10.1109/ACII.2009.5349477.

[23] Bousmalis, K., Mehu, M. and Pantic, M. 2013. Towards the Automatic Detection of Spontaneous Agreement and Disagreement Based on Nonverbal Behaviour: A Survey of Related Cues, Databases, and Tools. *Image Vision Comput.* 31, 2 (Feb. 2013), 203–221. DOI:https://doi.org/10.1016/j.imavis.2012.07.003.

[24] Bousmalis, K., Mehu, M. and Pantic, M. 2013. Towards the automatic detection of spontaneous agreement and disagreement based on nonverbal behaviour: A survey of related cues, databases, and tools. *Image and Vision Computing*. 31, 2 (2013), 203–221. DOI:https://doi.org/10.1016/j.imavis.2012.07.003.

[25] Bradley, B.H., Klotz, A.C., Postlethwaite, B.E. and Brown, K.G. 2013. Ready to rumble: How team personality composition and task conflict interact to improve performance. *Journal of Applied Psychology*. 98, 2 (2013), 385–392. DOI:https://doi.org/10.1037/a0029845.

[26] Bradley, J., Benyon, D., Mival, O. and Webb, N. 2010. Wizard of Oz experiments and companion dialogues. *Proceedings of the 24th BCS Interaction Specialist Group Conference* (Dundee, United Kingdom, Sep. 2010), 117–123.

[27] Bradley, M.M. and Lang, P.J. 1994. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*. 25, 1 (Mar. 1994), 49–59. DOI:https://doi.org/10.1016/0005-7916(94)90063-9.

[28] Brochet, M.G. 1986. *Effective Moderation of Computer Conferences: Notes and Suggestions*.

[29] Burke, R.J. 1970. Methods of resolving superior-subordinate conflict: The constructive use of subordinate differences and disagreements. *Organizational Behavior and Human Performance*. 5, 4 (Jul. 1970), 393–411. DOI:https://doi.org/10.1016/0030-5073(70)90029-2.

[30] Callaway, M.R., Marriott, R.G. and Esser, J.K. 1985. Effects of dominance on group decision making: Toward a stress-reduction explanation of groupthink. *Journal of Personality and Social Psychology*. 49, 4 (1985), 949–952. DOI:https://doi.org/10.1037/0022-3514.49.4.949.

[31] Carnevale, P.J. and Probst, T.M. 1998. Social values and social conflict in creative problem solving and categorization. *Journal of Personality and Social Psychology*. 74, 5 (1998), 1300–1309. DOI:https://doi.org/10.1037/0022-3514.74.5.1300.

[32] Cassell, J. 2001. Embodied Conversational Agents: Representation and Intelligence in User Interfaces. *AI Magazine*. 22, 4 (Dec. 2001), 67–67. DOI:https://doi.org/10.1609/aimag.v22i4.1593.

[33] Cassell, J. and Bickmore, T. 2000. External manifestations of trustworthiness in the interface. *Communications of the ACM*. 43, 12 (Dec. 2000), 50–56. DOI:https://doi.org/10.1145/355112.355123.

[34] Cassell, J., Sullivan, J., Churchill, E. and Prevost, S. 2000. *Embodied Conversational Agents*. MIT Press.

[35] Cassell, J., Vilhjálmsson, H.H. and Bickmore, T. 2004. BEAT: the Behavior Expression Animation Toolkit. *Life-Like Characters*. H. Prendinger and M. Ishizuka, eds. Springer Berlin Heidelberg. 163–185.

[36] Chen, H., Finin, T., Anupam Joshi, Kagal, L., Perich, F. and Dipanjan Chakraborty 2004. Intelligent agents meet the semantic Web in smart spaces. *IEEE Internet Computing*. 8, 6 (Nov. 2004), 69–79. DOI:https://doi.org/10.1109/MIC.2004.66.

[37] Chen, H., Finin, T., Joshi, A., Kagal, L., Perich, F. and Chakraborty, D. 2004. Intelligent agents meet the semantic Web in smart spaces. *IEEE Internet Computing*. 8, 6 (Nov. 2004), 69–79. DOI:https://doi.org/10.1109/MIC.2004.66.

[38] Chen, L., Harper, M., Franklin, A., Rose, T.R., Kimbara, I., Huang, Z. and Quek, F. 2006. A Multimodal Analysis of Floor Control in Meetings. *Machine Learning for Multimodal Interaction* (May 2006), 36–49.

[39] Chen, L. and Harper, M.P. 2009. Multimodal Floor Control Shift Detection. *Proceedings of the 2009 International Conference on Multimodal Interfaces* (New York, NY, USA, 2009), 15–22.

[40] Chen, M., Nikolaidis, S., Soh, H., Hsu, D. and Srinivasa, S. 2018. Trust-Aware Decision Making for Human-Robot Collaboration: Model Learning and Planning. *arXiv:1801.04099 [cs]*. (Nov. 2018).

[41] Christakopoulou, K., Radlinski, F. and Hofmann, K. 2016. Towards Conversational Recommender Systems. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (San Francisco California USA, Aug. 2016), 815–824.

[42] Conflict and well-being at work: the moderating role of personality | Journal of Managerial Psychology | Vol 20, No 2: *https://www.emeraldinsight.com/doi/full/10.1108/02683940510579740*. Accessed: 2018-03-23.

[43] Corbin, J.M., Strauss, A.L. and Strauss, A.L. 2008. *Basics of qualitative research: techniques and procedures for developing grounded theory*. Sage Publications, Inc.

[44] Cornelius, T.L., Alessi, G. and Shorey, R.C. 2007. The Effectiveness of Communication Skills Training With Married Couples: Does the Issue Discussed Matter? *The Family Journal*. 15, 2 (Apr. 2007), 124–132. DOI:https://doi.org/10.1177/1066480706297971.

[45] Correia, F., Alves-Oliveira, P., Maia, N., Ribeiro, T., Petisca, S., Melo, F. and Paiva, A. 2016. Just follow the suit! Trust in Human-Robot Interactions during Card Game Playing. (Aug. 2016).

[46] Correia, F., Mascarenhas, S., Prada, R., Melo, F.S. and Paiva, A. 2018. Group-based Emotions in Teams of Humans and Robots. *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (Chicago, IL, USA, Feb. 2018), 261–269.

[47] Correia, F., Petisca, S., Alves-Oliveira, P., Ribeiro, T., Melo, F. and Paiva, A. 2017. Groups of humans and robots: Understanding membership preferences and team formation. (Jul. 2017).

[48] Curşeu, P.L., Boroş, S. and Oerlemans, L.A.G. 2012. Task and relationship conflict in short-term and long-term groups: The critical role of emotion regulation. *International Journal of Conflict Management*. 23, 1 (Feb. 2012), 97–107. DOI:https://doi.org/10.1108/10444061211199331.

[49] Dahlbäck, N., Jönsson, A. and Ahrenberg, L. 1993. Wizard of Oz studies — why and how. *Knowledge-Based Systems*. 6, 4 (Dec. 1993), 258–266. DOI:https://doi.org/10.1016/0950-7051(93)90017-N.

[50] Dai Hasegawa, Justine Cassell and Kenji Araki 2010. The Role of Embodiment and Perspective in Direction-Giving Systems | Semantic Scholar. *AAAI Fall Symposium: Dialog with Robots* (2010).

[51] DeChurch, L.A. and Marks, M.A. 2001. Maximizing the benefits of task conflict: the role of conflict management. *International Journal of Conflict Management*. 12, 1 (Jan. 2001), 4–22. DOI:https://doi.org/10.1108/eb022847.

[52] Dehn, D.M. and Van mulken, S. 2000. The impact of animated interface agents: a review of empirical research. *International Journal of Human-Computer Studies*. 52, 1 (Jan. 2000), 1–22. DOI:https://doi.org/10.1006/ijhc.1999.0325.

[53] DeSanctis, G. and Brent Gallupe, R. 1987. A Foundation For The Study Of Group Decision Support Systems. *Management Science*. 33, (May 1987), 589–609. DOI:https://doi.org/10.1287/mnsc.33.5.589.

[54] DeSanctis, G. and Gallupe, R.B. 1987. A Foundation for the Study of Group Decision Support Systems. *Management Science*. 33, 5 (May 1987), 589–609. DOI:https://doi.org/10.1287/mnsc.33.5.589.

[55] Desert Survival Team Building Simulation: *https://www.humansynergistics.com/change-solutions/change-solutions-for-groups-and-teams/team-building-simulations/desert-survival-situation*. Accessed: 2020-05-09.

[56] Developing Facilitation Skills: *https://ctb.ku.edu/en/table-of-contents/leadership/group-facilitation/facilitation-skills/main*. Accessed: 2020-05-08.

[57] Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv:1810.04805 [cs]*. (May 2019).

[58] Diftler, M.A., Mehling, J.S., Abdallah, M.E., Radford, N.A., Bridgwater, L.B., Sanders, A.M., Askew, R.S., Linn, D.M., Yamokoski, J.D., Permenter, F.A., Hargrave, B.K., Platt, R., Savely, R.T. and Ambrose, R.O. 2011. Robonaut 2 - The first humanoid robot in space. *2011 IEEE International Conference on Robotics and Automation* (May 2011), 2178–2183.

[59] Douglas, T. 1970. *A Decade of Small Group Theory*. Bookstall Publications.

[60] Duysburgh, P., Elprama, S.A. and Jacobs, A. 2014. Exploring the Social-technological Gap in Telesurgery: Collaboration Within Distributed or Teams. *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (New York, NY, USA, 2014), 1537–1548.

[61] Farnham, S., Chesley, H.R., McGhee, D.E., Kawal, R. and Landau, J. 2000. Structured Online Interactions: Improving the Decision-making of Small Discussion Groups. *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work* (New York, NY, USA, 2000), 299–308.

[62] Farnham, S., Chesley, H.R., McGhee, D.E., Kawal, R. and Landau, J. 2000. Structured Online Interactions: Improving the Decision-making of Small Discussion Groups. *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work* (New York, NY, USA, 2000), 299–308.

[63] Fassaert, T., van Dulmen, S., Schellevis, F. and Bensing, J. 2007. Active listening in medical consultations: development of the Active Listening Observation Scale (ALOS-global). *Patient Education and Counseling*. 68, 3 (Nov. 2007), 258–264. DOI:https://doi.org/10.1016/j.pec.2007.06.011.

[64] Fischer-Lokou, J., Lamy, L., Guéguen, N. and Dubarry, A. 2016. Effects of Active Listening, Reformulation, and Imitation on Mediator Success: Preliminary Results. *Psychological Reports*. 118, 3 (Jun. 2016), 994–1010. DOI:https://doi.org/10.1177/0033294116646159.

[65] Freudian Defense Mechanisms and Empirical Findings in Modern Social Psychology: Reaction Formation, Projection, Displacement, Undoing, Isolation, Sublimation, and Denial - Baumeister - 1998 - Journal of Personality - Wiley Online Library: *https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-6494.00043?casa_token=irX9l-RDItUAAAAA:Kl7mkJnSC1ZurE-I9SikI75cjNVkharlEh2qKMxakTP4JIfJfKde8KCC4BKjK_NnnwUaVEUnDjnAjHA*. Accessed: 2020-06-02.

[66] Functional Roles of Group Members - Benne - 1948 - Journal of Social Issues - Wiley Online Library: *https://spssi.onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-4560.1948.tb01783.x*. Accessed: 2018-03-23.

[67] Galley, M., McKeown, K., Hirschberg, J. and Shriberg, E. 2004. Identifying agreement and disagreement in conversational speech: use of Bayesian networks to Model Pragmatic Dependencies. *Proceedings of the 42nd Meeting Association for Computational Linguistics*. (2004), 669–676.

[68] Gatica-Perez, D., McCowan, L., Zhang, D. and Bengio, S. 2005. Detecting Group Interest-Level in Meetings. *Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.* (Mar. 2005), 489–492.

[69] gensim: topic modelling for humans: *https://radimrehurek.com/gensim/models/doc2vec.html*. Accessed: 2020-05-09.

[70] Germesin, S. and Wilson, T. 2009. Agreement Detection in Multiparty Conversation. *Proceedings of the 2009 International Conference on Multimodal Interfaces* (New York, NY, USA, 2009), 7–14.

[71] Germesin, S. and Wilson, T. 2009. Agreement detection in multiparty conversation. *Proceedings of the 2009 international conference on Multimodal interfaces*. Id (2009), 7–14. DOI:https://doi.org/10.1145/1647314.1647319.

[72] Gers, F.A., Schmidhuber, J. and Cummins, F. 1999. Learning to forget: continual prediction with LSTM. (Jan. 1999), 850–855. DOI:https://doi.org/10.1049/cp:19991218.

[73] Gersick, C.J.G. 1988. Time and Transition in Work Teams: Toward a New Model of Group Development. *The Academy of Management Journal*. 31, 1 (1988), 9–41. DOI:https://doi.org/10.2307/256496.

[74] Graesser, A.C., Chipman, P., Haynes, B.C. and Olney, A. 2005. AutoTutor: an intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions on Education*. 48, 4 (Nov. 2005), 612–618. DOI:https://doi.org/10.1109/TE.2005.856149.

[75] Gratch, J., Wang, N., Gerten, J., Fast, E. and Duffy, R. 2007. Creating Rapport with Virtual Agents. *Intelligent Virtual Agents*. C. Pelachaud, J.-C. Martin, E. André, G. Chollet, K. Karpouzis, and D. Pelé, eds. Springer Berlin Heidelberg. 125–138.

[76] Greer, L.L. and Dannals, J.E. 2017. Conflict in Teams. *The Wiley Blackwell Handbook of the Psychology of Team Working and Collaborative Processes*. E. Salas, R. Rico, and J. Passmore, eds. John Wiley & Sons, Ltd. 317–343.

[77] Groom, V. and Nass, C. 2007. Can robots be teammates?: Benchmarks in human–robot teams. *Interaction Studies*. 8, 3 (Jan. 2007), 483–500. DOI:https://doi.org/10.1075/is.8.3.10gro.

[78] Groom, V. and Nass, C. 2007. Can robots be teammates? Benchmarks in human-robot teams. *Interaction Studies*. 8, (Oct. 2007), 483–500. DOI:https://doi.org/10.1075/is.8.3.10gro.

[79] Hacker, K.A., Amare, Y., Strunk, N. and Horst, L. 2000. Listening to youth: teen perspectives on pregnancy prevention. *Journal of Adolescent Health*. 26, 4 (Apr. 2000), 279–288. DOI:https://doi.org/10.1016/S1054-139X(99)00110-X.

[80] Hackman, J.R. and Katz, N. 2018. *Group Behavior and Performance*.

[81] Hahn, S. 2000. Agreement / Disagreement Classification : Exploiting Unlabeled Data using Contrast Classifiers. (2000).

[82] Hart, P. 1991. Irving L. Janis' Victims of Groupthink. *Political Psychology*. 12, 2 (1991), 247–278. DOI:https://doi.org/10.2307/3791464.

[83] Häuslschmid, R., von Bülow, M., Pfleging, B. and Butz, A. 2017. SupportingTrust in Autonomous Driving. *Proceedings of the 22nd International Conference on Intelligent User Interfaces* (Limassol, Cyprus, Mar. 2017), 319–329.

[84] Hayes, J. WORKPLACE CONFLICTAND HOW BUSINESSES CAN HARNESS IT TO THRIVE. 36.

[85] Heron, J. 1993. Group Facilitation: Theories and Models for Practice. (1993).

[86] Hill, G.W. 1982. Group versus individual performance: Are N + 1 heads better than one? (1982).

[87] Hillard, D., Ostendorf, M. and Shriberg, E. 2003. Detection Of Agreement vs. Disagreement In Meetings: Training With Unlabeled Data. *Companion Volume of the Proceedings of HLT-NAACL 2003 - Short Papers* (2003), 34–36.

[88] Horvath, A.O. and Greenberg, L.S. 1989. Development and validation of the Working Alliance Inventory. *Journal of Counseling Psychology*. 36, 2 (1989), 223–233. DOI:https://doi.org/10.1037/0022-0167.36.2.223.

[89] How Much Time Do We Spend in Meetings? (Hint: It's Scary): 2014. *https://www.themuse.com/advice/how-much-time-do-we-spend-in-meetings-hint-its-scary*. Accessed: 2019-05-13.

[90] Hybrid Forms of Third-Party Dispute Resolution: Theoretical Implications of Combining Mediation and Arbitration | Academy of Management Review: *https://journals.aom.org/doi/abs/10.5465/amr.2000.3312927*. Accessed: 2020-05-09.

[91] Ishii, R., Kumano, S. and Otsuka, K. 2015. Multimodal Fusion Using Respiration and Gaze for Predicting Next Speaker in Multi-Party Meetings. *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction* (New York, NY, USA, 2015), 99–106.

[92] Jackson, P. 1998. *Introduction to Expert Systems*. Addison-Wesley Longman Publishing Co., Inc.

[93] Jehn, K.A. 1995. A Multimethod Examination of the Benefits and Detriments of Intragroup Conflict. *Administrative Science Quarterly*. 40, 2 (1995), 256–282. DOI:https://doi.org/10.2307/2393638.

[94] Jehn, K.A. 1997. A Qualitative Analysis of Conflict Types and Dimensions in Organizational Groups. *Administrative Science Quarterly*. 42, 3 (1997), 530–557. DOI:https://doi.org/10.2307/2393737.

[95] Jehn, K.A. and Mannix, E.A. 2001. The Dynamic Nature of Conflict: A Longitudinal Study of Intragroup Conflict and Group Performance. *Academy of Management Journal*. 44, 2 (Apr. 2001), 238–251. DOI:https://doi.org/10.2307/3069453.

[96] Jehn, K.A. and Mannix, E.A. 2001. The Dynamic Nature of Conflict: A Longitudinal Study of Intragroup Conflict and Group Performance. *The Academy of Management Journal*. 44, 2 (2001), 238–251. DOI:https://doi.org/10.2307/3069453.

[97] Jordan, P.J. and Troth, A.C. 2004. Managing Emotions During Team Problem Solving: Emotional Intelligence and Conflict Resolution. *Human Performance*. 17, 2 (Apr. 2004), 195–218. DOI:https://doi.org/10.1207/s15327043hup1702_4.

[98] Jung, M.F. 2017. Affective Grounding in Human-Robot Interaction. *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (New York, NY, USA, 2017), 263–273.

[99] Jung, M.F. 2017. Affective Grounding in Human-Robot Interaction. *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (New York, NY, USA, 2017), 263–273.

[100] Jung, M.F., Martelaro, N. and Hinds, P.J. 2015. Using Robots to Moderate Team Conflict: The Case of Repairing Violations. *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (New York, NY, USA, 2015), 229–236.

[101] Jung, M.F., Martelaro, N. and Hinds, P.J. 2015. Using Robots to Moderate Team Conflict: The Case of Repairing Violations. *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (New York, NY, USA, 2015), 229–236.

[102] Jung, M.F., Murphy, R., Beane, M., Vertesi, J. and Forlizzi, J. 2017. Robots in Group Context : Rethinking Design , Development and Deployment. *Chi'17*. (2017), 1283–1288. DOI:https://doi.org/10.1145/3027063.3051136.

[103] Kahai, S.S., Sosik, J.J. and Avolio, B.J. 1997. Effects of Leadership Style and Problem Structure on Work Group Process and Outcomes in an Electronic Meeting System Environment. *Personnel Psychology*. 50, 1 (1997), 121–146. DOI:https://doi.org/10.1111/j.1744-6570.1997.tb00903.x.

[104] Kaner, S., Lind, L., Toldi, C., Fisk, S. and Berger, D. 1998. *Facilitator's Guide to Participatory Decision-Making*. New Society Publishers.

[105] Karacapilidis, N.I. and Pappis, C.P. 1997. A framework for group decision support systems: Combining AI tools and OR techniques. *European Journal of Operational Research*. 103, 2 (1997), 373–388. DOI:https://doi.org/Doi 10.1016/S0377-2217(97)00126-4.

[106] Karau, S.J. and Williams, K.D. 1993. Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology*. 65, 4 (1993), 681–706. DOI:https://doi.org/10.1037/0022-3514.65.4.681.

[107] Kim, B. and Rudin, C. 2013. Learning About Meetings. *arXiv:1306.1927 [cs, stat]*. (Jun. 2013).

[108] Kim, B. and Rudin, C. 2014. Learning about meetings. *Data Mining and Knowledge Discovery*. 28, 5–6 (2014), 1134–1157. DOI:https://doi.org/10.1007/s10618-014-0348-z.

[109] Kim, S., Valente, F. and Vinciarelli, A. 2012. Automatic detection of conflicts in spoken conversations: Ratings and analysis of broadcast political debates. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*. (2012), 5089–5092. DOI:https://doi.org/10.1109/ICASSP.2012.6289065.

[110] Kim, T., Chang, A., Holland, L. and Pentland, A.S. 2008. Meeting Mediator: Enhancing Group Collaborationusing Sociometric Feedback. *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work* (New York, NY, USA, 2008), 457–466.

[111] Kim, T., Chang, A., Holland, L. and Pentland, A.S. 2008. Meeting Mediator: Enhancing Group Collaborationusing Sociometric Feedback. *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work* (New York, NY, USA, 2008), 457–466.

[112] Koda, T. and Maes, P. 1996. Agents with faces: the effect of personification. *Proceedings 5th IEEE International Workshop on Robot and Human Communication. RO-MAN'96 TSUKUBA* (Nov. 1996), 189–194.

[113] Kourmousi, N., Amanaki, E., Tzavara, C. and Koutras, V. 2017. Active Listening Attitude Scale (ALAS): Reliability and Validity in a Nationwide Sample of Greek Educators. *Social Sciences*. 6, 1 (2017), 1–14.

[114] Kuhn T and Poole MS 2006. Do conflict management styles affect group decision making? Evidence from a longitudinal field study. *Human Communication Research*. 26, 4 (Jan. 2006), 558–590. DOI:https://doi.org/10.1111/j.1468-2958.2000.tb00769.x.

[115] Kumar, R. and Rosé, C.P. 2011. Architecture for Building Conversational Agents that Support Collaborative Learning. *IEEE Transactions on Learning Technologies*. 4, 1 (Jan. 2011), 21–34. DOI:https://doi.org/10.1109/TLT.2010.41.

[116] Latan?, B. 1981. The psychology of social impact. *American Psychologist*. 36, 4 (1981), 343–356. DOI:https://doi.org/10.1037/0003-066X.36.4.343.

[117] Lawrence, P.R. and Lorsch, J.W. 1967. Differentiation and Integration in Complex Organizations. *Administrative Science Quarterly*. 12, 1 (1967), 1–47. DOI:https://doi.org/10.2307/2391211.

[118] Lee, T., Chang, C. and Chen, G. 2007. Building an Interactive Caring Agent for Students in Computer-based Learning Environments. *Seventh IEEE International Conference on Advanced Learning Technologies (ICALT 2007)* (Jul. 2007), 300–304.

[119] Lester, J.C., Converse, S.A., Kahler, S.E., Barlow, S.T., Stone, B.A. and Bhogal, R.S. 1997. The persona effect: affective impact of animated pedagogical agents. *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems* (Atlanta, Georgia, USA, Mar. 1997), 359–366.

[120] Leyzberg, D., Spaulding, S., Toneva, M. and Scassellati, B. 2012. The Physical Presence of a Robot Tutor Increases Cognitive Learning Gains. *Proceedings of the Annual Meeting of the Cognitive Science Society*. 34, 34 (2012).

[121] Luger, E. and Sellen, A. 2016. "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA, May 2016), 5286–5297.

[122] Luhtanen, R. and Crocker, J. 1992. A Collective Self-Esteem Scale: Self-Evaluation of One's Social Identity. *Personality and Social Psychology Bulletin*. 18, 3 (Jun. 1992), 302–318. DOI:https://doi.org/10.1177/0146167292183006.

[123] Malcolm C. Burson Spring 20. Finding Clarity in the Midst of Conflict:Facilitating Dialogue and Skillful DiscussionUsing a Model from the Quaker Tradition. *Group Facilitation: A Research and Applications Journ*. 4, (Spring 20).

[124] Matsuyama, Y., Akiba, I., Fujie, S. and Kobayashi, T. 2015. Four-participant group conversation: A facilitation robot controlling engagement density as the fourth participant. *Computer Speech & Language*. 33, 1 (Sep. 2015), 1–24. DOI:https://doi.org/10.1016/j.csl.2014.12.001.

[125] Matsuyama, Y., Akiba, I., Fujie, S. and Kobayashi, T. 2015. Four-participant group conversation: A facilitation robot controlling engagement density as the fourth participant. *Computer Speech & Language*. 33, (2015), 1–24. DOI:https://doi.org/10.1016/j.csl.2014.12.001.

[126] Matsuyama, Y., Akiba, I., Saito, A. and Kobayashi, T. 2013. A four-participant group facilitation framework for conversational robots. *SIGDIAL 2013 - 14th Annual Meeting of the Special Interest Group on Discourse and Dialogue, Proceedings of the Conference* (2013).

[127] Matsuyama, Y. and Kobayashi, T. 2015. Towards a Computational Model of Small Group Facilitation. *2015 AAAI Spring Symposium Series* (Mar. 2015).

[128] McGregor, M. and Tang, J.C. 2017. More to Meetings: Challenges in Using Speech-Based Technology to Support Meetings. *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA, Feb. 2017), 2208–2220.

[129] McMillan, S.S., King, M. and Tully, M.P. 2016. How to use the nominal group and Delphi techniques. *International Journal of Clinical Pharmacy*. 38, (2016), 655–662. DOI:https://doi.org/10.1007/s11096-016-0257-x.

[130] Michinov, N. and Michinov, E. 2008. Face-to-face contact at the midpoint of an online collaboration: Its impact on the patterns of participation, interaction, affect, and behavior over time. *Computers & Education*. 50, 4 (May 2008), 1540–1557. DOI:https://doi.org/10.1016/j.compedu.2007.03.002.

[131] Mikic, I., Huang, K. and Trivedi, M. 2000. Activity monitoring and summarization for an intelligent meeting room. (2000), 107–112.

[132] Mikolov, T., Karafiat, M., Burget, L., Cernocky, J. and Khudanpur, S. Recurrent Neural Network Based Language Model. 4.

[133] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. and Dean, J. 2013. Distributed Representations of Words and Phrases and their Compositionality. *arXiv:1310.4546 [cs, stat]*. (Oct. 2013).

[134] Miranda, S.M. and Bostrom, R.P. 1999. Meeting Facilitation: Process Versus Content Interventions. *Journal of Management Information Systems*. 15, 4 (Mar. 1999), 89–114. DOI:https://doi.org/10.1080/07421222.1999.11518223.

[135] Miranda, S.M. and Bostrom, R.P. 1997. Meeting facilitation: process versus content interventions. *Proceedings of the Thirtieth Hawaii International Conference on System Sciences* (Jan. 1997), 124–133 vol.2.

[136] Moreno, R., Mayer, R.E. and Lester, J.C. 2000. Life-Like Pedagogical Agents in Constructivist Multimedia Environments: Cognitive Consequences of their Interaction. (2000).

[137] Moreno, R., Mayer, R.E., Spires, H.A. and Lester, J.C. 2001. The Case for Social Agency in Computer-Based Teaching: Do Students Learn More Deeply When They Interact With Animated Pedagogical Agents? (2001). DOI:https://doi.org/10.1207/S1532690XCI1902_02.

[138] Moubayed, S.A. and Skantze, G. 2011. Turn-taking Control Using Gaze in Multiparty Human-Computer Dialogue: Effects of 2D and 3D Displays. (2011), 4.

[139] Mulken, S. van, André, E. and Müller, J. 1999. An empirical study on the trustworthiness of life-like interface agents. *Proceedings of the HCI International '99 (the 8th International Conference on Human-Computer Interaction) on Human-Computer Interaction: Communication, Cooperation, and Application Design-Volume 2 - Volume 2* (USA, Aug. 1999), 152–156.

[140] Murphy, R.R. 2004. Human-robot interaction in rescue robotics. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*. 34, 2 (May 2004), 138–153. DOI:https://doi.org/10.1109/TSMCC.2004.826267.

[141] Nass, C., Fogg, B.J. and Moon, Y. 1996. Can computers be teammates? *International Journal of Human-Computer Studies*. 45, 6 (1996), 669–678. DOI:https://doi.org/10.1006/ijhc.1996.0073.

[142] Neil Katz, Kevin McNult, D.A. 1994. Conflict Resolution. (1994), 19.

[143] Niederman, F. and Volkema, R.J. 1999. The Effects of Facilitator Characteristics on Meeting Preparation, Set Up, and Implementation. *Small Group Research*. 30, 3 (Jun. 1999), 330–360. DOI:https://doi.org/10.1177/104649649903000304.

[144] Niederman, F. and Volkema, R.J. 1999. The Effects of Facilitator Characteristics on Meeting Preparation, Set Up, and Implementation. *Small Group Research*. 30, 3 (Jun. 1999), 330–360. DOI:https://doi.org/10.1177/104649649903000304.

[145] Noesner, G.W. and Webster, M. 1997. Crisis Intervention: Using Active Listening Skills in Negotiations. *FBI Law Enforcement Bulletin*. 66, (1997), 13.

[146] Novick, D. and Gris, I. 2014. Building Rapport between Human and ECA: A Pilot Study. *Human-Computer Interaction. Advanced Interaction Modalities and Techniques* (2014), 472–480.

[147] Nowak, K.L. and Biocca, F. 2003. The Effect of the Agency and Anthropomorphism on Users' Sense of Telepresence, Copresence, and Social Presence in Virtual Environments. *Presence*. 12, 5 (Oct. 2003), 481–494. DOI:https://doi.org/10.1162/105474603322761289.

[148] Nunamaker, J.F., Briggs, R.O., Mittleman, D.D., Vogel, D.R. and Balthazard, P.A. 1996. Lessons from a Dozen Years of Group Support Systems Research: A Discussion of Lab and Field Findings. *J. of Management Information Systems*. 13, (1996), 163–207. DOI:https://doi.org/10.1080/07421222.1996.11518138.

[149] Nunamaker, J.F., Jr., Briggs, R.O., Mittleman, D.D., Vogel, D.R. and Balthazard, P.A. 1996. Lessons from a Dozen Years of Group Support Systems Research: A Discussion of Lab and Field Findings. *J. Manage. Inf. Syst.* 13, 3 (Dec. 1996), 163–207. DOI:https://doi.org/10.1080/07421222.1996.11518138.

[150] NUNAMAKER, JR, J.F. 1997. Future research in group support systems: needs, some questions and possible directions. *International Journal of Human-Computer Studies*. 47, 3 (1997), 357–385. DOI:https://doi.org/10.1006/ijhc.1997.0142.

[151] Panagakis, Y., Zafeiriou, S. and Pantic, M. 2015. Audiovisual Conflict Detection in Political Debates. *Computer Vision - ECCV 2014 Workshops* (Cham, 2015), 306–314.

[152] Panko, R.R. and Kinney, S.T. 1995. Meeting profiles: size, duration, and location. *Proceedings of the Twenty-Eighth Annual Hawaii International Conference on System Sciences* (Jan. 1995), 1002–1011 vol.4.

[153] Paper, C. 2017. Robot Chameleons and Small Group Decision Making : The Case of Conformity. July (2017).

[154] Pennington, J., Socher, R. and Manning, C. 2014. Glove: Global Vectors for Word Representation. (2014), 1532–1543.

[155] Perlow, L.A., Hadley, C.N. and Eun, E. 2017. Stop the Meeting Madness. *Harvard Business Review*.

[156] Phillips, B. 1999. Reformulating dispute narratives through active listening. *Mediation Quarterly*. 17, 2 (1999), 161–180. DOI:https://doi.org/10.1002/crq.3890170207.

[157] Poole, M.S., Holmes, M. and DeSanctis, G. 1991. Conflict management in a computer-supported meeting environment. *Management Science*. 37, 8 (Aug. 1991), 926–953.

[158] Poole, M.S., Homes, M. and DeSanctis, G. 1988. Conflict management and group decision support systems. (1988), 227–243.

[159] Porcheron, M., Fischer, J.E. and Sharples, S. 2016. Using Mobile Phones in Pub Talk. *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (San Francisco, California, USA, Feb. 2016), 1649–1661.

[160] Rahim, M.A., Psenicka, C., Polychroniou, P., Zhao, J.-H., Yu, C.-S., Chan, K.A., Susana, K.W.Y., Alves, M.G., Lee, C.-W., Ralunan, S., Ferdausy, S. and Wyk, R. van 2002. A model of emotional intelligence and conflict management strategies: a study in seven countries. *The International Journal of Organizational Analysis*. 10, 4 (Apr. 2002), 302–326. DOI:https://doi.org/10.1108/eb028955.

[161] Riek, L.D. 2012. Wizard of Oz studies in HRI: a systematic review and new reporting guidelines. *Journal of Human-Robot Interaction*. 1, 1 (Jul. 2012), 119–136. DOI:https://doi.org/10.5898/JHRI.1.1.Riek.

[162] Rienks, R. and Verbree, D. 2005. Twente Argument Schema Annotation Manual v 0.99b. (2005), 22.

[163] Ripple effects of an embedded social agent | Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: *https://dl.acm.org/doi/abs/10.1145/2207676.2207776*. Accessed: 2020-05-08.

[164] Romano, N.C. and Nunamaker, J.F. 2001. Meeting analysis: findings from research and practice. *Proceedings of the 34th Annual Hawaii International Conference on System Sciences* (Jan. 2001), 13 pp.-.

[165] Schuman, S.P. 1996. The Role of Facilitation in Collaborative Groups. *Creating Collaborative Advantage*. SAGE Publications Ltd. 127–140.

[166] Shamekhi, A., Liao, Q.V., Wang, D., Bellamy, R.K.E. and Erickson, T. 2018. Face Value? Exploring the Effects of Embodiment for a Group Facilitation Agent. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2018), 391:1–391:13.

[167] Shen, S., Slovak, P. and 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 - HRI '18*. (2018), 69–77. DOI:https://doi.org/10.1145/3171221.3171248.

[168] Short, E. and Matari, M.J. Towards Autonomous Moderation of an Assembly Game.

[169] Short, E. and Mataric, M.J. 2017. Robot moderation of a collaborative game: Towards socially assistive robotics in group interactions. *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (Aug. 2017), 385–390.

[170] Skantze, G. and Al Moubayed, S. 2012. IrisTK: A statechart-based toolkit for multi-party face-to-face interaction. (Sep. 2012).

[171] Skantze, G., Johansson, M. and Beskow, J. 2015. Exploring Turn-taking Cues in Multi-party Human-robot Discussions About Objects. *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction* (New York, NY, USA, 2015), 67–74.

[172] Small-Group Research in Social Psychology: Topics and Trends over Time: *https://www.slideshare.net/marina761/smallgroup-research-in-social-psychology-topics-and-trends-over-time*. Accessed: 2018-03-23.

[173] Strohkorb Sebo, S., Traeger, M., Jung, M. and Scassellati, B. 2018. The Ripple Effects of Vulnerability. *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction - HRI '18*. (2018), 178–186. DOI:https://doi.org/10.1145/3171221.3171275.

[174] Syna Desivilya, H. and Yagil, D. 2004. *The Role of Emotions in Conflict Management: The Case of Work Teams*. Technical Report #ID 602041. Social Science Research Network.

[175] Takayama, L. and Go, J. 2012. Mixing metaphors in mobile remote presence. *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work* (Seattle, Washington, USA, Feb. 2012), 495–504.

[176] Takeuchi, A. and Naito, T. 1995. Situated facial displays: towards social interaction. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA, May 1995), 450–455.

[177] Team, 3M Meeting Management and Drew, J. 1994. *Mastering Meetings: Discovering the Hidden Potential of Effective Business Meetings*. McGraw-Hill.

[178] Tekleab, A.G., Quigley, N.R. and Tesluk, P.E. 2009. A Longitudinal Study of Team Conflict, Conflict Management, Cohesion, and Team Effectiveness. *Group & Organization Management*. 34, 2 (Apr. 2009), 170–205. DOI:https://doi.org/10.1177/1059601108331218.

[179] Tennent, H., Shen, S. and Jung, M. 2019. Micbot: A Peripheral Robotic Object to Shape Conversational Dynamics and Team Performance. *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Mar. 2019), 133–142.

[180] The Group Affect and Performance (GAP) Corpus | Proceedings of the Group Interaction Frontiers in Technology: *https://dl.acm.org/doi/abs/10.1145/3279981.3279985*. Accessed: 2020-05-09.

[181] The Uncanny Valley: The Original Essay by Masahiro Mori - IEEE Spectrum: 1970. *https://spectrum.ieee.org/automaton/robotics/humanoids/the-uncanny-valley*. Accessed: 2020-06-28.

[182] Thomas, K. and Kilmann, R. 1974. Thomas-Kilmann Conflict Style Questionnaire. (1974).

[183] Thornton, R.K. and Sokoloff, D.R. 1998. Assessing student learning of Newton's laws: The Force and Motion Conceptual Evaluation and the Evaluation of Active Learning Laboratory and Lecture Curricula. (1998).

[184] Time, Interaction, and Performance (TIP): A Theory of Groups - Joseph E. Mcgrath, 1991: *https://journals.sagepub.com/doi/10.1177/1046496491222001*. Accessed: 2020-06-28.

[185] Tuckman, B.W. 1965. Developmental sequence in small groups. *Psychological Bulletin*. 63, 6 (1965), 384–399. DOI:https://doi.org/10.1037/h0022100.

[186] Tuckman, B.W. and Jensen, M.A.C. 1977. Stages of Small-Group Development Revisited. *Group & Organization Studies*. 2, 4 (Dec. 1977), 419–427. DOI:https://doi.org/10.1177/105960117700200404.

[187] Turban, E., Aronson, J.E., Liang, T.-P. and Sharda, R. 2006. *Decision Support and Business Intelligence Systems*. Prentice Hall.

[188] Viller, S. 1991. The Group Facilitator: A CSCW Perspective. *Proceedings of the Second European Conference on Computer-Supported Cooperative Work ECSCW '91*. Springer, Dordrecht. 81–95.

[189] Viller, S. 1991. The Group Facilitator: A CSCW Perspective. *Proceedings of the Second European Conference on Computer-Supported Cooperative Work ECSCW '91*. L. Bannon, M. Robinson, and K. Schmidt, eds. Springer Netherlands. 81–95.

[190] Waibel, A., Schultz, T., Bett, M., Denecke, M., Malkin, R., Rogina, I., Stiefelhagen, R. and Jie Yang 2003. SMaRT: the Smart Meeting Room Task at ISL. *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03).* (Apr. 2003), IV–752.

[191] Waibel, A., Schultz, T., Bett, M., Denecke, M., Malkin, R., Rogina, I., Stiefelhagen, R. and Yang, J. 2003. SMaRT: the Smart Meeting Room Task at ISL. *2003 IEEE International Conference on*

Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03) (Apr. 2003), IV-752–5 vol.4.

[192] Walker, J.H., Sproull, L. and Subramani, R. 1994. Using a human face in an interface. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, Massachusetts, USA, Apr. 1994), 85–91.

[193] Wang, L. and Cardie, C. 2016. Improving Agreement and Disagreement Identification in Online Discussions with A Socially-Tuned Sentiment Lexicon. (2016). DOI:https://doi.org/10.3115/v1/W14-2617.

[194] Wang, W., Yaman, S., Precoda, K., Richey, C., Raymond, G., International, S.R.I., Avenue, R. and Park, M. 2011. Detection of Agreement and Disagreement in Broadcast Conversations. *Computational Linguistics*. (2011), 374–378.

[195] Watson, D., Clark, L.A. and Tellegen, A. 1988. Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology*. 54, 6 (Jun. 1988), 1063–1070.

[196] Westley, F. and Waters, J.A. 1988. Group Facilitation Skills for Managers. *Management Education and Development*. 19, 2 (Jun. 1988), 134–143. DOI:https://doi.org/10.1177/135050768801900207.

[197] Westley, F. and Waters, J.A. 1988. Group Facilitation Skills for Managers. *Management Education and Development*. 19, 2 (Jun. 1988), 134–143. DOI:https://doi.org/10.1177/135050768801900207.

[198] de Wit, F.R.C., Greer, L.L. and Jehn, K.A. 2012. The paradox of intragroup conflict: a meta-analysis. *The Journal of Applied Psychology*. 97, 2 (Mar. 2012), 360–390. DOI:https://doi.org/10.1037/a0024844.

[199] de Wit, F.R.C., Jehn, K.A. and Scheepers, D. 2013. Task conflict, information processing, and decision-making: The damaging effect of relationship conflict. *Organizational Behavior and Human Decision Processes*. 122, 2 (Nov. 2013), 177–189. DOI:https://doi.org/10.1016/j.obhdp.2013.07.002.

[200] de Wit, F.R.C., Jehn, K.A. and Scheepers, D. 2013. Task conflict, information processing, and decision-making: The damaging effect of relationship conflict. *Organizational Behavior and Human Decision Processes*. 122, 2 (Nov. 2013), 177–189. DOI:https://doi.org/10.1016/j.obhdp.2013.07.002.

[201] Yang, J. and Mossholder, K.W. 2004. Decoupling task and relationship conflict: the role of intragroup emotional processing. *Journal of Organizational Behavior*. 25, 5 (Aug. 2004), 589–605. DOI:https://doi.org/10.1002/job.258.

[202] Yu, Z. and Nakamura, Y. 2010. Smart Meeting Systems: A Survey of State-of-the-art and Open Issues. *ACM Comput. Surv.* 42, 2 (Mar. 2010), 8:1–8:20. DOI:https://doi.org/10.1145/1667062.1667065.

Appendix A: Study Questionnaires

Sociodemographic Questionnaire

Please take a moment and answer a few questions about yourself:

Date of Birth: _________

Gender: M / F / Prefer to self-describe______________________________

Ethnic Background (check all that apply):

American Indian or Alaskan Native ____

Asian ____

Native Hawaiian or other Pacific Islander ____

Black or African American ____

White ____

Hispanic or Latino ____

Prefer to self-describe _________________________________

Last grade of school completed (check one):

Less than high school (0-8) ____

Some high school ____

High school graduate or GED____

Technical school education ____

Some college ____

College graduate ____

Advanced degree ____

Occupation: ____

How often do you read books (check one):

Never ____

Less than once a week ____

Once a week ____

A few times a week ____

Every day ____

How much experience do you have with <u>computers</u> (check one):

I've never used one. ____

I've tried one a few times. ____

I use one regularly. ____

I'm an expert. ____

How much experience do you have with <u>robots</u> (check one):

I've never used one. ____

I've tried one a few times. ____

I use one regularly. ____

I'm an expert. ____

How much experience do you have with <u>conversational agents (e.g., Siri, Alexa, Google home)</u> (check one):

I've never used one. ____

I've tried one a few times. ____

I use one regularly. ____

I’m an expert. ____

How do you feel about using computers (check one):

I don’t like them. ____

They’re OK. ____

They can be useful. ____

I love playing with them. ____

How much experience do you have attending meetings in order to arrive at a group consensus on a decision (group decision making)?

I’ve never been in a group decision-making meeting. ____

I’ve been in group decision-making meetings a few times ____

I regularly attend group decision-making meetings as a participant ____

I regularly lead group decision-making meetings ____

Ten-Item Personality Inventory-(TIPI)

Here are a number of personality traits that may or may not apply to you. Please write a number next to each statement to indicate the extent to which you agree or disagree with that statement. You should rate the extent to which the pair of traits applies to you, even if one characteristic applies more strongly than the other.

Disagree strongly	Disagree moderately	Disagree a little	Neither agree nor disagree	Agree a little	Agree moderately	Agree strongly
1	2	3	4	5	6	7

I see myself as:

1. _____ Extraverted, enthusiastic.

2. _____ Critical, quarrelsome.

3. _____ Dependable, self-disciplined.

4. _____ Anxious, easily upset.

5. _____ Open to new experiences, complex.

6. _____ Reserved, quiet.

7. _____ Sympathetic, warm.

8. _____ Disorganized, careless.

9. _____ Calm, emotionally stable.

10. _____ Conventional, uncreative.

TIPI scale scoring (‘‘R’’ denotes reverse-scored items): Extraversion: 1, 6R; Agreeableness: 2R, 7; Conscientiousness; 3, 8R; Emotional Stability: 4R, 9; Openness to Experiences: 5, 10R

Leadership skills questionnaire

Directions 1. Place yourself in the role of a leader when responding to this questionnaire.

2. Respond to each of the statements below using the following scale:

Not true	Seldom true	Occasionally true	Somewhat true	Very true
1	2	3	4	5

1. I am effective with the detailed aspects of my work.
2. I usually know ahead of time how people will respond to a new idea or proposal.
3. I am effective at problem solving.
4. Understanding the social fabric of the organization is important to me.
5. When problems arise, I immediately address them.
6. Managing people and resources is one of my strengths.
7. I am able to sense the emotional undercurrents in my group.
8. Seeing the big picture comes easily for me.
9. In my work, I enjoy responding to people's requests and concerns.
10. I use my emotional energy to motivate others.
11. The key to successful conflict resolution is respecting my opponent.
12. I enjoy discussing organizational values and philosophy.
13. I am effective at obtaining resources to support our programs.
14. I work hard to find consensus in conflict situations.
15. I am flexible about making changes in our organization.

Feelings Towards Group Work

This questionnaire contains a number of statements about how you might feel about working in groups. If you feel the statement is very true of you, circle the “5”. If you feel the statement is not true of you at all, circle the “1” If you feel the statement is partly true of you, circle the “2”, “3” or “4”. Remember that there are no right or wrong answers. Please respond to all statements.

Never true of me	Seldom true	Occasionally true	Somewhat true	Always true of me
1	2	3	4	5

1. I enjoy working within a group (I) *

2. I sometimes feel nervous when I have to give my ideas or communicate within a group (D)

3. I often find it difficult to understand what the group task is (D)

4. I like to work alone even when placed in a group (I)

5. I think groups should take the time to set up group goals

6. I prefer to work within a group rather than work alone (I) *

7. Even when the group is achieving its goals, I don’t really feel involved or satisfied

8. I often have a strong feeling satisfaction when I become totally involved in a group achievement (G)

9. I don’t like it when one member of the group takes over from everyone else

10. Groups should organize themselves so that the work is divided evenly

11. I usually make a strong personal contribution to group work (G)

12. I like group work more when we can make up our own groups (G).

13. Contributing ideas within a group often makes me feel better about myself

14. I can usually understand other group members’ ideas (G)

15. Even when groups are well organized, I don’t believe they are a more effective way of using time (I)

16. It is best when each person helps each other within a group (G)

17. I often think the work becomes too confusing when done in a group rather than individually (I)

28. I rarely feel relaxed within a group (D)

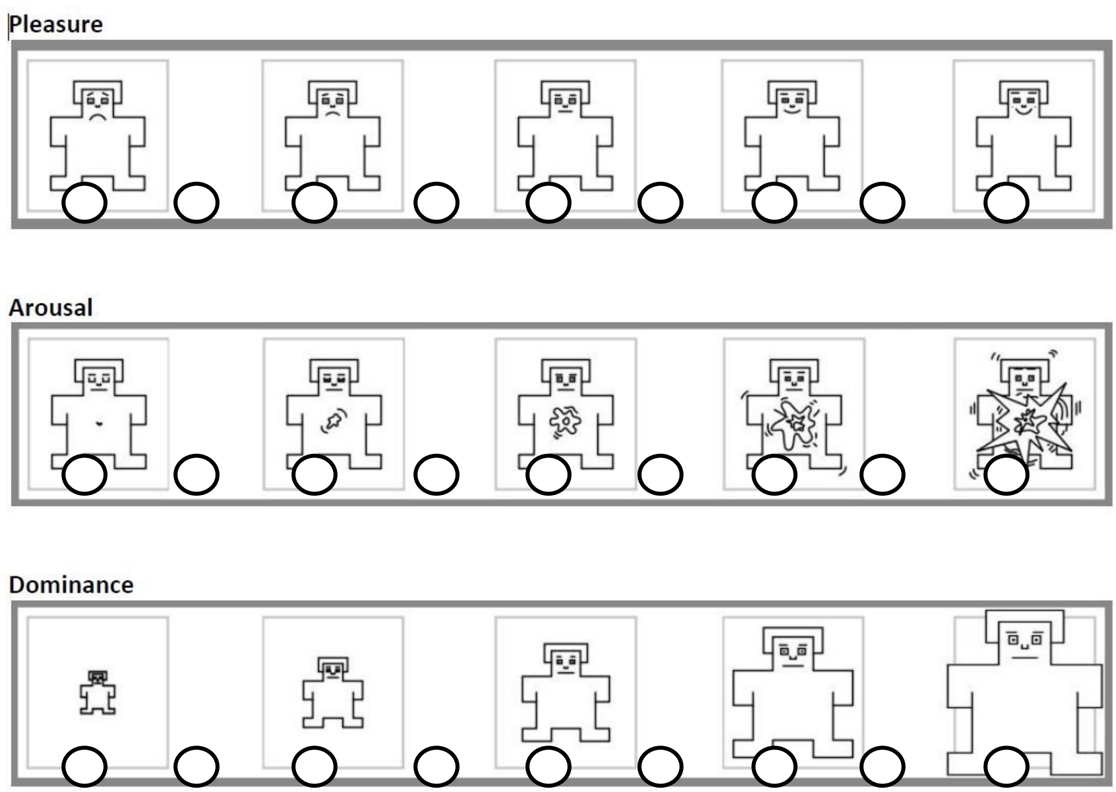
19. I sometimes feel let down by other group members (I)

20. I often feel in charge when working within a group

SAM Scales

The participants self-reported their momentary feelings of pleasure (valance), arousal and dominance using a validated 9-point pictorial rating scale (the Self-Assessment Manikin or SAM: Bradley & Lang, 1994)

Direction: Please place an ‘x’ over any of the five figures in each scale, or between any two figures.



Positive and Negative Affect Schedule (PANAS)

This scale consists of a number of words that describe different feelings and emotions. Read each item and then select the appropriate answer from the list below. Indicate to what extent you feel this way right now, that is, at the present moment.

NOTE: Pick one of the following and write to the right of each word:

very slightly or not at all	a little	moderately	quite a bit	extremely
1	2	3	4	5

interested	irritable
distressed	alert
excited	ashamed
upset	inspired
strong	nervous
guilty	determined
scared	attentive
hostile	jittery
enthusiastic	active
proud	afraid

Intra Group Conflict

Never	Rarely	Occasionally	Sometimes	Frequently	Usually	Always
1	2	3	4	5	6	7

Relationship conflict:

How frequently . . .

- . . . was there friction among members on your team?
- . . . were personality conflicts evident on your team?
- . . . was there tension among members on your team?
- . . . was there emotional conflict among members on your team?

Task conflict:

How frequently . . .

- . . . did people on your team disagree regarding the work being done?
- . . . were there conflicts about ideas on your team?
- . . . were there differences of opinion on your team?

Conflict management:

To what extent do you agree with the following statements?

- Conflict is dealt with openly on this team.
- If conflict arises on this team, the people involved in the conflict initiate steps to resolve the conflict immediately.
- This team knows what to do when conflicts between team members arise.
- This team is able to avoid the negative aspects of conflict before they occur.

Cohesion, Performance, Efficiency and Dominance

Disagree strongly	Disagree moderately	Disagree a little	Neither agree nor disagree	Agree a little	Agree moderately	Agree strongly
1	2	3	4	5	6	7

Cohesion: To what extent do you agree with the following statements?

- Our team is united in trying to reach its goals for performance.
- We all take responsibility for any loss or poor performance by our team.
- Our team members communicate freely about each of our personal responsibilities in getting this project done.
- The members of this team help each other when working on our project.
- The members of this team get along well together.
- The members of this team stick together.

Satisfaction: To what extent do you agree with the following statements?

- I am satisfied with our decision-making **process** in the meeting.
- I am happy with our decision **outcome**.
- I am pleased with the way my partner and I **participated** in the decision-making task.

Perceived performance:

Please indicate a score (out of 100) your group performance deserves to earn ________.

Perceived efficiency:

Please indicate a score (out of 100) your group efficiency deserves to earn ________.

Dominance:

- From 1 to 7, how would you rate **your** level of dominance in the group discussion you just experienced ________.
- From 1 to 7, how would you rate **your partner's** level of dominance in the group discussion you just experienced ________.

Automated Group Facilitator Evaluation Questionnaire 1

Please answer the following questions about the facilitation agent:

Disagree strongly	Disagree moderately	Disagree a little	Neither agree nor disagree	Agree a little	Agree moderately	Agree strongly
1	2	3	4	5	6	7

1. I am satisfied with the interaction with the facilitator.
2. I found **it easy to interact** with the facilitator.
3. I would use the facilitation agent **again** for group decision-making tasks.
4. I felt that I could **trust** the facilitation agent leaders during today's session.
5. I found the facilitation agent **helpful** for our decision-making process. [I am confident in the group leaders' ability to help me.]
6. I found the facilitation agent **knowledgeable**.
7. I felt the facilitation agent was powerful and confident.
8. I felt the facilitation agent was **friendly** and warm toward me.
9. The facilitation agent was able to manage the discussion and interaction during the meeting.
10. The facilitation agent attempted to manage the conflicts raised in the meeting.
11. The facilitation agent could effectively manage the disagreements between my partner and I.
12. The facilitation agent assured that my partner and I have equal chance to express our ideas.
13. The facilitator allowed sufficient discussion.
14. The facilitator encouraged participation
15. The facilitator helped bring out new group ideas.
16. The facilitator helped close out discussions.

Automated Group Facilitator Evaluation Questionnaire 2

In this part, we ask you to report your general impression of the agent, by rating on pairs of adjectives describing the agent. Please pay attention to the pair of adjectives in each question.

Adjective A						**Adjective B**
1	**2**	**3**	**4**	**5**	**6**	**7**

In general, I found the agent to be: *

Adjective A: Fake	Adjective B: Natural
Adjective A: Machine-like	Adjective B: Natural
Adjective A: Unconscious	Adjective B: Natural
Adjective A: Dislikable	Adjective B: Natural
Adjective A: Unfriendly	Adjective B: Natural
Adjective A: Unpleasant	Adjective B: Natural
Adjective A: Incompetent	Adjective B: Natural
Adjective A: Ignorant	Adjective B: Natural
Adjective A: Unintelligent	Adjective B: Natural
Adjective A: Weak	Adjective B: Natural
Adjective A: Submissive	Adjective B: Natural
Adjective A: Lacking confidence	Adjective B: Natural
Adjective A: Unreliable	Adjective B: Natural
Adjective A: Untrustworthy	Adjective B: Natural
Adjective A: Insincere	Adjective B: Natural

Rapport with the Facilitation Agent

Please answer the following questions about the facilitation agent:

Disagree strongly	Disagree moderately	Disagree a little	Neither agree nor disagree	Agree a little	Agree moderately	Agree strongly
1	2	3	4	5	6	7

1. The agent seemed engaged in our discussions.
2. I felt I had a connection with the agent.
3. I felt the agent and I understood each other.
4. I felt like the agent was NOT paying attention to what I said.
5. I felt the agent and I were working towards a common goal.
6. I made efforts to respond to the agent's questions and suggestions.

Working Alliance Inventory (Short Form)

Instructions: On the following pages there are sentences that describe some of the different ways a person might think or feel about his or her facilitator. As you read the sentences mentally insert the name of your facilitator in place of _____________in the text.

Below each statement inside there is a seven-point scale:

1	2	3	4	5	6	7
Never	Rarely	Occasionally	Sometimes	Often	Very Often	Always

If the statement describes the way you ***always*** feel (or think) circle the number 7; if it ***never*** applies to you circle the number 1. Use the numbers in between to describe the variations between these extremes.

1._______________ and I agree about the steps to be taken to improve his/her situation.

2. The facilitator and I both feel confident about the usefulness of our current activity in therapy.

3.I believe _______________ likes me.

4.I have doubts about what we are trying to accomplish in the session.

5. I am confident in my ability to help _______________.

6. We are working towards mutually agreed upon goals.

7. I appreciate _______________ as a person.

8. We agree on what is important for _______________ to work on.

9._______________ and I have built a mutual trust.

10._______________ and I have different ideas on what his/her real problems are.

11.We have established a good understanding between us of the kind of changes that would be good for _______________.

12._______________ believes the way we are working with her/his problem is correct.

Team/partner Evaluation Questionnaire (Satisfaction with the team)

Please indicate to what extent you agree with the following statements about your team and your partner?

Disagree strongly	Disagree moderately	Disagree a little	Neither agree nor disagree	Agree a little	Agree moderately	Agree strongly
1	2	3	4	5	6	7

1. I am satisfied with my team members.
2. I am pleased with the way my team members and I worked together.
3. I am very satisfied with working in this team.
4. I found it pleasant interacting with my partner.
5. I liked my partner.
6. I found my partner and I shared many similarities.
7. My partner seemed engaged in our discussions.
8. I felt my partner and I understood each other.
9. I felt like my partner was NOT paying attention to what I said.
10. I felt my partner and I were working towards a common goal.
11. I made efforts to respond to my partner's questions and suggestions.

Meeting Evaluation **Questionnaire**

Please indicate to what extent you agree with the following statements about your group decision-making meeting?

Disagree strongly	Disagree moderately	Disagree a little	Neither agree nor disagree	Agree a little	Agree moderately	Agree strongly
1	2	3	4	5	6	7

2. I found the decision-making task to be difficult.
3. I am NOT happy with the result of our discussions.
4. I found my partner and I worked together successfully.
5. We achieved the planned objectives of the meeting.
6. In general, I liked our discussions for the decision.
7. I found it difficult to bring my ideas into the discussion.
8. I found we reached a decision efficiently.
9. My partner and I were given equal time to express our ideas.
10. The meeting time was managed wisely.
11. The meeting was well-planned.

Semi structured interview questions

1. What were your overall impressions of the study?
2. How did you feel about the facilitation agent?
3. Do you feel she was helpful in:
    a. Improving the meeting structure?
    b. Managing conflicts?
    c. Reducing your biases?

4. How would you describe the group decision making process and experience?
5. What was the best thing about the facilitation agent you noticed?
6. What were the worst things the facilitation agent you noticed?
7. Do you have any suggestions to improve this system?
8. Would you use the automated group facilitation system in the future?
9. Would you recommend using automated group facilitation system to your friends?
10. Is there anything else you would like to add?

Group Facilitation Robot Evaluation

Please answer the following questions about the facilitation agent:

Disagree strongly	Disagree moderately	Disagree a little	Neither agree nor disagree	Agree a little	Agree moderately	Agree strongly
1	2	3	4	5	6	7

17. I am satisfied with the interaction with the robot facilitator.
18. I found **it easy to interact** with the robot facilitator.
19. I would use the facilitation robot **again** for group decision-making tasks.
20. I felt that I could **trust** the facilitation robot leaders during today's session.
21. I found the facilitation robot **helpful** for our decision-making process. [I am confident in the robot facilitator's ability to help me.]
22. I found the facilitation robot **knowledgeable**.
23. I felt the facilitation robot was powerful and confident.
24. I felt the facilitation robot was **friendly** and warm toward me.
25. The facilitation robot was able to manage the discussion and interaction during the meeting.
26. The facilitation robot attempted to manage the conflicts raised in the meeting.
27. The facilitation robot could effectively manage the disagreements between my partner and I.
28. The facilitation robot assured that my partner and I have equal chance to express our ideas.
29. The robot facilitator allowed sufficient discussion.
30. The robot facilitator encouraged participation
31. The robot facilitator helped bring out new group ideas.
32. The robot facilitator helped close out discussions.
33. The robot facilitator helped managing the meeting structure and the conversation flow.
34. The robot facilitator was effective in handling the disagreement between the members of our group.
35. The robot facilitator was effective in balancing the participation of all members of the group.
36. The facilitator was helpful in managing the time of the meeting.
37. The facilitator provided useful information about the alternative options.

Team Efficacy[10]

1. Achieving this team's goals is well within our reach.

2. This team can achieve its task without requiring us to put in unreasonable time or effort.

3. With focus and effort, this team can do anything we set out to accomplish.

Team psychological safety[2]

(7-point scale from "very inaccurate" to "very accurate.)

Disagree strongly	Disagree moderately	Disagree a little	Neither agree nor disagree	Agree a little	Agree moderately	Agree strongly
1	2	3	4	5	6	7

1. If you make a mistake on this team, it is often held against you.

2. Members of this team are able to bring up problems and tough issues.

3. People on this team sometimes reject others for being different.

4. It is safe to take a risk on this team.

5. It is difficult to ask other members of this team for help.

6. No one on this team would deliberately act in a way that undermines my efforts.

7. Working with members of this team, my unique skills and talents are valued and utilized

Dominance:

- From 1 to 7, how would you rate **your** level of dominance in the group discussion you just experienced ________.

---

[10] Edmondson, A., 1999. Psychological safety and learning behavior in work teams. *Administrative science quarterly*, *44*(2), pp.350-383.

- From 1 to 7, how would you rate **your partner's** level of dominance in the group discussion you just experienced ________.

(7-point Likert scale)

I feel all the group members collaborated equally to make the final decision.

I feel I was given the change to express all my ideas.

Conflict management:

My partners and I had (small) disagreement on the decision-making task but we could resolve them smoothly.

The robot helped me and my partners to smoothly resolve our disagreements on the decision-making task.

Assessment of Contribution

(on scale of 1 - 100)

1. Please rate to what degree your teammate and the robot contributed to the **decision making** in your team? **(from 1 to 100)**
   - Your teammate contribution in the decision-making: ____________
   - The robot (Sarah) contribution in the decision-making: ____________

2. Please rate to what degree your teammate and the robot contributed to the **meeting management** (e.g. managing the structure and the time during the meeting) in your team? **(from 1 to 100)**
   - Your teammate contribution in the meeting management: ____________
   - The robot (Sarah) contribution in the meeting management: ____________

3. Please rate to what degree the robot actively contributed to the **conflict/disagreement management** in your team? **(from 1 to 100)**
   - The robot (Sarah) contribution in the conflict management: ____________

Task and Conflict recall

1. How many items did you discuss in your team in total? __________
2. How many <u>times</u> did you and your teammate have disagreements?

   Step 1(pair comparison): _________ Step2(ranking): _________
3. How many <u>times</u> did you and your teammate have complete agreement?

   Step 1(pair comparison): _________ Step2(ranking): _________

Attitude towards Team Conflict

Please indicate to what extent you agree with the following statements:

Disagree strongly	Disagree moderately	Disagree a little	Neither agree nor disagree	Agree a little	Agree moderately	Agree strongly
1	**2**	**3**	**4**	**5**	**6**	**7**

1. I believe conflicts/disagreements among members of a group are totally acceptable.

   _____
2. I believe conflicts/disagreements among members of a group are always manageable.

   _____
3. I believe conflicts/disagreements among members of a group are always resolvable.

   _____

Group Atmosphere

Please answer the following questions:

1. How much do you **trust** your fellow group members?

Not at all			Neutral			I completely trust him/her
1	**2**	**3**	**4**	**5**	**6**	**7**

2. How **comfortable** do you feel **delegating** to your group members?

Not comfortable at all			Neutral			Comfortable A lot
1	2	3	4	5	6	7

3. Were your group members truthful and honest?

Not at all			Neutral			s/he was 100% truthful and honest
1	2	3	4	5	6	7

4. How much do you **respect** your fellow group members?

Not at all			Neutral			Very much
1	2	3	4	5	6	7

5. How much do you **respect the ideas** of the people in your group?

Not at all			Neutral			Very much
1	2	3	4	5	6	7

6. How much do you **like** your group members?

Not at all			Neutral			I Like him/her A lot
1	2	3	4	5	6	7

7. To what degree was **communication** in your group **open**?

Not open at all			Neutral			Completely open
1	2	3	4	5	6	7

8. To what degree was **conflict dealt with openly** in your work group?

Not at all			Neutral			Completely openly
1	2	3	4	5	6	7

9. To what extent is your group **cohesive**?

Not at all			Neutral			Very much Cohesive
1	2	3	4	5	6	7

10. How much do you feel like your team has **group spirit**?

Not at all			Neutral			Very much
1	2	3	4	5	6	7

11. To what degree would you consider this group as a **great group** to work in?

Not at all			Neutral			Very great group (best)
1	2	3	4	5	6	7

Meeting Evaluation Questionnaire

Please indicate to what extent you agree with the following statements about your group decision-making meeting?

Disagree strongly	Disagree moderately	Disagree a little	Neither agree nor disagree	Agree a little	Agree moderately	Agree strongly
1	**2**	**3**	**4**	**5**	**6**	**7**

1. This task took longer than expected to complete. _______

2. Our group worked well together. _______

3. Our group used its time wisely. _______

4. Our group struggled to work together efficiently on this task. _______

5. Overall, our group did a good job on this task. _______

6. I helped lead the group during this task. _______

Active Listening Attitude Scale (ALAS):

Please indicate the extent to which you agree with the following statements about yourself and your teammate during the group activity you just had.

Disagree strongly	Disagree moderately	Disagree a little	Neither agree nor disagree	Agree a little	Agree moderately	Agree strongly
1	**2**	**3**	**4**	**5**	**6**	**7**

Statement about ME	Agreement score 1-7	**Statement about my TEAMMATE**	Agreement score 1-7
I listened to my teammate calmly, while he/she was speaking.		My teammate listened to me calmly, while I was speaking.	
I listened to my teammate, putting myself in his/her shoes.		My teammate listened to me, putting herself/himself in my shoes.	
I sometimes gave my teammate a brief summary of what he/she has said.		My teammate sometimes gave me a brief summary of what I have said.	
I kept tracks of points my teammate made.		My teammate kept tracks of points I made.	

Sometimes I began to talk before my teammate finishes talking.		My teammate sometimes began to talk before I finish talking.	
I assured my teammate that I was listening by using verbal acknowledgement.		My teammate assured me that he/she was listening by using verbal acknowledgement.	
I assured my teammate that I was receptive to him/her ideas.		My teammate assured me that he/she was receptive to my ideas.	
I showed my teammate that I was listening by my body language (e.g. head nods).		My teammate showed me that he/she was listening by his/her body language (e.g. head nods).	
I began arguing with the other person before I know it, while I was listening to him/her.		My teammate began arguing with me before he/she knows it, while he/she was listening to me.	
When I want to say something, I talk about it, even if I interrupt the other person.		When my teammate wants to say something, she/he talks about it, even if she/he interrupts me.	
I tended to deny the other person's opinion, when it's different from mine.		My teammate tended to deny my opinion, when it's different from his/hers.	
I tended to persist in my opinion, while talking with my teammate.		My teammate tended to persist in his/her opinion, while talking with me.	

Appendix B. Study Protocols

B. Study 2

SMALL GROUP DECISION MAKING FACILITATION

STUDY PROTOCOL

(1/8/2019)

1. **[Randomize the group into one of 2 arms before starting. Set up the study room, microphone on and on the table, Job descriptions on the table, Furhat is on. Run the server. open questionnaire tabs on Firefox on tablets. PRINT consent and debriefing forms. Prepare study money]**

2. **[Give each group a unique ID. Use group's unique ID as pre-fix to participant's ID (e.g: G01P1, for Group ID 1 Partner ID 1 for questionnaires)]**

3. Thank you for coming in today to help us out with our research. Let me first just double check that you are eligible. Are you both at least 21 years old? And able to speak and read English? Have you two had some professional experience, including co-ops, internships, or full-time employment? **[If no, thank and dismiss, if yes continue.]**

4. Let me tell you a little bit about what we will be doing in this study. We're developing [a robot/technologies] that will help small groups improve their performance, and meeting quality during a group decision making task and we want to see how well our system works. In this study we will be simulating a hiring decision making meeting. You will be given a set of resumes and will be asked to review and discuss them to make a GROUP decision and select the one best candidate collaboratively [**robot:** and a robot will be supporting your meeting]. In addition to [talking to the robot/doing the group decision making task], we will also ask you to fill out some questionnaires before and after [talking to the robot/the meeting]. The study session today should take about one and half hour and, at the end, you will be compensated $20 each. Does this sound like something you are still interested in? **[If yes, continue, if no, dismiss]**

5. Ok. Great. Now the first thing to do will be reading and signing the consent form. I will explain it to you and you can take as much time as you want to read it, and then ask any questions you may have. **[Administer informed consent]**

    a. 21 years old, speak and read English
    b. Explanation (you already said this)

c. Takes one session, about 1.5 hour.
d. No risks, can skip any question on a questionnaire you're not comfortable with
e. Information is confidential (stored in locked cabinet, locked room), nothing linked to your name
f. STUDY IS COMPLETELY VOLUNTARY, stop any time
g. $20 for completing study
h. Do you have any further questions?

6. Now, I would like you to fill out several questionnaires

7. Here are the questionnaires. Some of these provide background information about yourself. Some ask you questions about your group behaviors and how you are feeling right now. **[Hand participants tablets with the pre questionnaires window open (firefox browser): Socio-Demographic + Personality Inventory (TIPI), + Leadership Skills, + Feeling towards group work, + Self-Assessment Manikin (SAM), + Positive And Negative Affect Schedule (PANAS)]**

8. Now I would like to tell you a little about the application you will be using in this study

   [Run the server and show the web application on the chrome browser]. Here is the web application that you will be using in this session. First each of you need to **login** with your name, your id and your group id [remind participants where they can find their id and group id]. You will use the app to review the resumes, rank them and discuss them. You will be given a job description printed on a paper and when the session starts, **[robot: at certain points the robot will tell you to]** use the application to review the resumes individually and rank them, so you use the app to do so [show them the initial ranking page, tap on a resume, show arrows and show the submit button and the <u>confidence question</u>, give them a minute to try it themselves]. In addition to your individual rating, when you start discussing the candidates together to make a group decision, the robot will ask you to use the application during your decision-making process, to indicate which candidate you are discussing and your decision. Most of your actions will be mirrored in your team-mate tablet [show the discussion page, tap on a resume, tap on discuss button, show eliminate and keep button, make sure they understand and give them 1-2 minutes to try].

9. [When they are done. **Turn on the videotape, Re-run the server**, **run the Furhat program (only for robot condition)**, and **login** the participants] Now I turn on the cameras and I am going to take you to the study room.

10. [Go to the other room. **Ask them not to sit** on the chairs until you introduce the agent.]

11. **a) For robot:** Ok, Here is our robot named Sarah. Like I said earlier, this robot will moderate/facilitate your hiring decision-making process in this session. It will give you instructions and guide you through different steps. When I start the robot, the robot will listen, talk to you and you can answer to the robot almost like the way you would talk to a human. There may be some delays time to time for her to understand your word.

b) For control: Ok, now we can get started. Here are some instructions for your decision-making process [Hand them the written instruction paper]. Like I said earlier, you will review the resumes individually first and then discuss together to select one best candidate. When you make a decision, you use the application again to enter your final individual rating again. You have about 30 minutes for this session. You can start now, and I will come back to the room in 30 minutes**.**

The whole session will be **videotaped**. If either of you becomes uncomfortable during the session, you may stop the session. Also, I will be monitoring the session from another room. If a health or safety concern arises, let me know and I will immediately call for the appropriate help.

12. Now I login you to in the application before you sit.

13. [Guide each participant to seat on their seat (participant ID 1 seats first and participant ID 2 seats after him/her.) Tell them they first need to stand until Furhat turns to them and nod her head. Then they can seat.]

14. **[Start the program from the third tablet (RA's tablet). When furhat starts speaking, step outside.]**

15. **a) For robot: [Robot Session terminates. Turn off the videotape. Re-enter the room, make sure participants submitted the final voting]**

    b) For control: [After 30 minutes reenter the room. And turn off the videotape, make sure participants submitted the final voting]

16. Great. Thank you so much for completing the group decision-making task. Now I just need you to fill out a few more questionnaires. Let me take you back to the previous room. **[Hand in the tablet with the post questionnaires window open (on Firefox browser): Self-Assessment Manikin (SAM), + Positive And Negative Affect Schedule (PANAS), + Intra Group Conflict + Group Performance, + Team Evaluation, + Meeting Evaluation, + <u>For Robot condition: Group Facilitator Evaluation, + Rapport with robot, + Working Alliance Inventory with robot</u>].**

17. Great. That concludes the computer portion of the study. Lastly, I would just like to ask you a few questions about your experience using the system in your group decision

making**.** Is it OK if I record the conversation? It is just so I can take detailed notes. **[Conduct semi-structured interview, use phone to record]**

18. **[Debrief participants (debrief statement)]**

19. Alright, now if you just sign this payment receipt you will be on your way. Thank you again for participating. **[Pay participant $20 each, have them sign receipt]**

20. **[Save the Database data before running the server again]**

B. Study 3

SMALL GROUP DECISION MAKING FACILITATION

STUDY PROTOCOL

(1/16/2020)

1. **[Randomize the group into one of 2 arms before starting. Set up the study room, microphone on and on the table, Task descriptions on the table, Furhat is on. Put the <u>participant names</u> in the program. Run the program. Start the java app. Open questionnaire tabs on Firefox on tablets. PRINT consent and debriefing forms. Prepare the study money]**

2. **[Give each group a unique ID. Use group's unique ID as pre-fix to participant's ID (e.g: G01P1, for Group ID 1 Partner ID 1 for questionnaires)]**

3. Thank you for coming in today to help us out with our research. Let me first just double check that you are eligible. Are you both at least 21 years old? And able to speak and read English? Have you two had some professional experience, including co-ops, internships, or full-time employment? **[If no, thank and dismiss, if yes continue.]**

4. Let me tell you a little bit about what we will be doing in this study. We're developing <u>[some technologies including a robot-driven system]</u> that will help small groups improve their performance, and meeting quality during a group decision making task and we want to see how well our system works [**robot:** The robot is especially designed to **moderate disagreements** among the members of the group].

In this study we will be simulating a "*Winter Survival Exercise*". You will be asked to imagine yourself in a hypothetical survival situation in which your plane crash landed in a very cold weather and you must collaborate as a team and **make multiple group decisions to survive** in a plane crashed situation. You will be given a task description displayed on a screen and all instructions are provided by a robot facilitator.

During this session you will first **make decisions individually**. Then you will be asked to review and discuss the decisions in your group to make GROUP decisions collaboratively and a robot will be supporting your meeting. After the group discussion you will be asked to indicate your decisions again individually.

During your group discussion, you may have disagreements on some items and it's totally fine. We actually want you to express and defend your opinions about each item. After your group discussion and final decision, we will share a standard solution for the winter survival situation. Your **group decisions will be evaluated** against this standard solution.

After this group session, **each of you will be evaluated individually** based on your contribution to the final group decisions. You will be each given 5 points for every decision you make that appears in the final group decision. So if you think a decision is important you should express your thoughts.

You have about **40 minutes** for this session to complete the task and come up with your final decision.

5. In addition to [talking to the robot/doing the survival exercise], we will also ask you to fill out some questionnaires before and after [talking to the robot/the task]. The study session today should take about one and half hour and, at the end, you will be compensated $20 each. Does this sound like something you are still interested in? **[If yes, continue, if no, dismiss]**

6. Ok. Great. Now the first thing to do will be reading and signing the consent form. I will explain it to you and you can take as much time as you want to read it, and then ask any questions you may have. **[Administer informed consent]**

    a. 21 years old, speak and read English
    b. Explanation (you already said this)
    c. Takes one session, about 1.5 hour.
    d. No risks, can skip any question on a questionnaire you're not comfortable with
    e. Information is confidential (stored in locked cabinet, locked room), nothing linked to your name
    f. STUDY IS COMPLETELY VOLUNTARY, stop any time
    g. $20 for completing study
    h. Do you have any further questions?

7. Now, I would like you to fill out several questionnaires

8. Here are the questionnaires. Some of these provide background information about yourself. Some ask you questions about your group behaviors and how you are feeling right now. **[Hand participants tablets with the pre questionnaires window open (firefox browser): Socio-Demographic + Personality Inventory (TIPI), + Leadership Skills, + Feeling towards group work, + Self-Assessment Manikin (SAM), + Positive And Negative Affect Schedule (PANAS)]**

9. **NEW:** Hand each participant a scenario page. Tell each subject unique information about the need for certain items to appear in the final selection. Tell each subject it is important for their items to appear in the final selection. Tell subjects the study is about assessing their persuasion and negotiation skills.

10. Now I would like to show you the form you will use to review and rank the items individually before and after your group discussion. You will be given the task description displayed on a shared screen and when the session starts, **[robot: at certain points the robot will tell you to]** use the tablet to review the items individually and rank them.

11. [When they are done. **Turn on the videotape**, **run the Furhat program (only for robot condition)**] Now I turn on the cameras and I am going to take you to the study room.

12. [Go to the other room. **Ask them not to sit** on the chairs until you introduce the robot.]

13. **a) For robot:** Ok, Here is our robot named Sarah. Like I said earlier, this robot will moderate/facilitate your decision-making process in this winter survival exercise session. It will give you instructions and guide you through different steps. When I start the robot, the robot will listen, talk to you and you can answer to the robot almost like the way you would talk to a human. There may be some delays time to time for her to understand your word.

    b) For control: Ok, now we can get started. Here are some instructions for your decision-making process [Hand them the written instruction paper]. Like I said earlier, you will review the resumes individually first and then discuss together to select one best candidate. When you make a decision, you use the application again to enter your final individual rating again. You have about 30 minutes for this session. You can start now, and I will come back to the room in 30 minutes**.**

The whole session will be **videotaped**. If either of you becomes uncomfortable during the session, you may stop the session. Also, I will be monitoring the session from another room. If a health or safety concern arises, let me know and I will immediately call for the appropriate help.

14. [Guide each participant to seat on their seat (participant ID 1 seats first and participant ID 2 seats after him/her.) Tell them they first need to stand until Furhat turns to them and nod her head. Then they can seat.]

15. **[Start the program from the third tablet (RA's tablet). When furhat starts speaking, step outside.]**

16. **a) For robot: [Robot Session terminates. Turn off the videotape. Re-enter the room, make sure participants submitted the final voting]**

    b) For control: [After 30 minutes reenter the room. And turn off the videotape, make sure participants submitted the final voting]

17. Great. Thank you so much for completing the group decision-making task. Now I just need you to fill out a few more questionnaires. Let me take you back to the previous room. **[Hand in the tablet with the post questionnaires window open (on Firefox browser): Self-Assessment Manikin (SAM), + Positive And Negative Affect Schedule (PANAS), + Intra Group Conflict + Group Performance, + Team Evaluation, + Meeting Evaluation, + <u>For Robot condition: Group Facilitator Evaluation, + Rapport with robot, + Working Alliance Inventory with robot</u>].**

18. Great. That concludes the computer portion of the study. Lastly, I would just like to ask you a few questions about your experience using the system in your group decision making**.** Is it OK if I record the conversation? It is just so I can take detailed notes. **[Conduct semi-structured interview, use phone to record]**

19. **[Debrief participants (debrief statement)]**

20. Alright, now if you just sign this payment receipt you will be on your way. Thank you again for participating. **[Pay participant $20 each, have them sign receipt]**

Semi structured interview questions

1. What were your overall impressions of the session?

a. How did you feel about your group decision making session?
b. How would you describe the group decision making process and experience?
c. Remember one of the latest group meetings that you have attended and compare it to that session.

2. How would you describe Sarah (the group facilitation system)? For example if you want to describe it to a friend.

3. How did you feel about the facilitation agent (facilitation system)?
    a. Probe: in what ways was it helpful?
    b. Probe: Improving the meeting structure?
    c. Probe: Balancing participation?
    d. Probe: Improving performance?

4. What did you like most about the group facilitation system?
5. What did you like the least about the group facilitation system?
6. Do you have any suggestions to improve this system?
7. Would you use the automated group facilitation system in the future? Why? How?
8. Would you recommend using automated group facilitation system to your friends?
9. Is there anything else you would like to add?
10. Control group – Vignette experiment: Ask participants to imagine a situation where they had disagreement and the robot intervene to manage their conflict. What would they think about that.

Appendix C. Written Instructions for Participants

C. Task Descriptions

Winter Survival Exercise

The Situation:

- You have just crash-landed in the woods of northern Minnesota and southern Manitoba. It is 11:32 A.M. in mid-January.
- The small plane in which you were traveling has been completely destroyed except for the frame. The pilot and co-pilot have been killed, but no one else is seriously injured.
- The crash came suddenly before the pilot had time to radio for help or inform anyone of your position. Since your pilot was trying to avoid a storm, you know the plane was considerably off course. The pilot announced shortly before the crash that you were eighty miles northwest of a small town that is the nearest known habitation.
- You are in a wilderness area made up of thick woods broken by many lakes and rivers.
- The last weather report indicated that the temperature would reach minus twenty-five degrees in the daytime and minus forty at night.
- You are dressed in winter clothing appropriate for city wear—suits, pantsuits, street shoes, and overcoats.
- While escaping from the plane, your group salvaged the **10** items listed below.

Your task is to rank these items according to their importance to your survival.

• You may assume that the amount of each item is the same as the number in your group and that the group has agreed to stick together.

Winter Survival Decision Items:

Rank the following items according to their importance to your survival, starting with "**1" for the most important** and proceeding to "10" for the least important!

- Compress kit (with 28 ft. of 2-inch gauze)
- Ball of steel wool
- Cigarette lighter without the fluid

- Loaded .45-caliber pistol
- Pocket Mirror
- 30 feet of rope
- Family-sized chocolate bar (one per person)
- Flashlight with batteries
- Quart of 85-proof whiskey
- Extra shirt and pants for each survivor

C. Instructions for Small Group Decision-Making Session

The goal of the meeting is to select an applicant to interview for a sales manager position in a company. Human resources has selected six resumes, out of 70 applicants, and now we would like you to help us select the **ONE** best candidate for interview. You have **30 minutes** to select the best candidate.

In the tablet in front of you, you can review the six resumes, and the job description is printed on the paper.

We suggest you follow the steps below during your decision-making process:

1- Get to know each other [Introduction and greeting].

2- Use your tablet to review and rank the resumes **individually** for about 5 minutes.

3- When you are done with you individual ranking, start discussing the important criteria in your group.

4- Then you may start your **group** discussion by reviewing the resumes one by one, and eliminate the ones that are clearly below the bar based on your criteria. You can use the tablets for reviewing resumes, and eliminating them.

5- When you narrow down your choices to 2 or 3 candidates, you can discuss the pros and cons of each of the remaining candidates and explain who and why you think is the best fit for this position.

6- Then you try to <u>**finalize your decision on the ONE best candidate**</u> to interview.

7- Finally please use the tablet to rank the resumes one more time after your discussion.

8- Call the research assistant when you are done.

Appendix D. Debriefing Documents

Debriefing Statement –Group Facilitation System – STUDY 3

The experiment that you just participated in was designed to investigate whether a conversational robot can facilitate a group decision-making session and manage the disagreements among the group members as well as the meeting structure and member participation.

There are three separate conditions in this study, and participants were randomly assigned to, and experienced one of them:

1. Robot Facilitation System with Disagreement Management Intervention: In this condition the group decision-making session is facilitated by a conversational robot. The robot's language and behaviors were controlled by a research assistant in an adjacent room who remote-controlled the robot. The facilitation robot provides resolution interventions when there is a disagreement between participants and attempts to manage the meeting by following the agenda and monitoring the members' speaking time and participation.

2. Robot Facilitation System without Disagreement Management Intervention: In this condition, the decision-making session is facilitated by a robot that is controlled by a research assistant from an adjacent room. However, the robot does not attempt to resolve disagreements between participants.

3. No Facilitation: In this condition participants are asked to make a group decision without any type of robot facilitation.

Please understand that this deception was an absolutely critical component of our research, and that this experiment was carefully evaluated and approved by the Northeastern University Institutional Review Board, which reviews all experiments that involve human subjects.

If at any time, now or later, you experience any ill effects, either mental or physical, as a result of your participation in this experiment, please do not hesitate to tell the experimenter, or to contact Dr. Timothy Bickmore at (617) 373-5477, or bickmore@ccs.neu.edu. You have the option to withdraw your data from this experiment now so that it will not be used, with no repercussions. If you choose to withdraw, no data or video record of your participation will be kept.

Please note that this is an ongoing study, so please do not disclose any information about the study conditions to your friends and classmates. Feel free to ask any questions about the experiment at this time.

Your help has been greatly appreciated, and will aid the Human-Computer Interaction Laboratory in the construction of new software that can facilitate group meetings and ultimately will be used to assist people in making group decisions more efficiently and saving a large amount of time and money.