MIA: Motivational Interviewing Agent for Improving Conversational Skills in Remote Group Discussions

SAMIHA SAMROSE, University of Rochester, USA

EHSAN HOQUE, University of Rochester, USA

Since online discussion platforms can limit the perception of social cues, effective collaboration over videochat requires additional attention to conversational skills. However, *self-affirmation* and *defensive bias* theories indicate that feedback may appear confrontational, especially when users are not motivated to incorporate them. We develop a feedback chatbot that employs Motivational Interviewing (MI), a directive counseling method that encourages commitment to behavior change, with the end goal of improving the user's conversational skills. We conduct a within-subject study with 21 participants in 8 teams to evaluate our MI-agent 'MIA' and a non-MI-agent 'Roboto'. After interacting with an agent, participants are tasked with conversing over videochat to evaluate candidate résumés for a job circular. Our quantitative evaluation shows that the MI-agent effectively motivates users, improves their conversational skills, and is likable. Through a qualitative lens, we present the strategies and the cautions needed to fulfill individual and team goals during group discussions. Our findings reveal the potential of the MI technique to improve collaboration and provide examples of conversational tactics important for optimal discussion outcomes.

CCS Concepts: • **Human-centered computing** → *Empirical studies in collaborative and social computing*.

Additional Key Words and Phrases: group; discussion; video chat; agent; feedback; motivational interviewing

1 INTRODUCTION

The success of a team can be measured from task performance and group satisfaction [76], both of which depend on maintaining the appropriate collaborative behaviors. Such behaviors can be difficult to maintain because people can often violate the norms without even realizing (e.g., shouting, over-participation)[21]. The challenges intensify even more when the discussion happens over remote mediums (e.g., videochat), where social cues are more difficult to notice [78]. The importance of improving group discussions over videochat has become prominent especially with the increased usage of videoconferencing platforms during the COVID-19 pandemic[1]. Feedback, reflections, and reminders have been found useful in conversational behavior modulations for videochat meetings [22, 32, 45, 63, 64].

---

[1]https://www.marketwatch.com/story/zoom-microsoft-cloud-usage-are-rocketing-during-coronavirus-pandemic-new-data-show-2020-03-30

Authors' addresses: Samiha Samrose, University of Rochester, USA, ssamrose@cs.rochester.edu; Ehsan Hoque, University of Rochester, USA, mehoque@cs.rochester.edu.

Even though feedback can positively impact collaborative behaviors for videochat meetings, there are two major research gaps. Firstly, feedback for improving group discussions often assumes that users already are or would be motivated to modify their behaviors as per the corresponding feedback. The application of motivation theories is often missing in such feedback. It is especially important because *self-affirmation* and *defensive bias* theories indicate that feedback may appear confrontational, and thus users may have internal resistance towards executing the feedback [69, 70]. Secondly, users often need regular and actionable suggestions that explain how a particular behavior feature can be improved [63]. *Self-reflection*-based feedback showcases how the users performed [41], whereas *suggestion-oriented* feedback recommends how the performance can be modified [14]. Developing and delivering such suggestive feedback as reminders through virtual agents before the videochat meeting also remains under-explored.

In this work, we explore how motivation theories can be incorporated in feedback agents in order to facilitate conversational behavior changes in remote group discussions. We adopt Motivational Interviewing (MI) [48] which is a directive counseling method for encouraging commitment towards behavior change. The approach has been found effective, especially in high-risk scenarios (e.g., moderating alcohol consumption or maintaining healthy diet [54, 67]), but low-risk scenarios such as group conversations may still be able to benefit from MI. Rather than incorporating MI into therapy, our focus is to incorporate the technique to the dialogue structure of a chatbot that helps the users understand and improve conversational skills. We build in suggestions to the agent which also works as a pre-meeting feedback reminders. Incorporating MI in a feedback agent is challenging, because MI evokes motivation by building up on the user's own viewpoints, whereas current virtual agents have limited capabilities in continuing an open-ended conversation. The conversational style of current virtual agents appears mostly transactional and closed-ended. Even though previous research has explored therapeutic and coaching agents which allow users to converse in a more *open-ended* context [2, 3, 28], it has not been explored intensively in MI-based agents. One reason is that MI is comparatively a new agent property, thus *close-ended* interaction allows more control over the human-machine interaction. Following this style, recent research with MI-agents successfully improved habits such as maintaining a healthy diet [54, 55]. Besides agent performance, understanding task performance in terms of what strategies users apply during the actual decision-making process can also be helpful to iterate on the suggestive feedback component of the agents.

In the first phase, we design a wireframe prototype of a chatbot with motivational interviewing properties as per literature review and suggestions by *the Center for Student Conflict Management* at University of Rochester. We build a non-MI agent 'Roboto' as a control case to compare the two agents' performances. The dialogue of 'Roboto' focuses on effective information delivery in an understandable manner, similar to previous chatbots in literature [65, 76]. We conduct a treatment-control-based within-subject user study with 21 participants in 8 teams, tasked with coming to a consensus during a discussion over video chat. For the discussion task, we prepare and rank some résumés for a given job circular with the help of *the Greene Center for Career Education & Connections* at University of Rochester. Each participant in a team is assigned an individual candidate résumé. In the group videochat discussion, the group goal is to unanimously elect a candidate for the job, whereas the individual goal is to convince other team members to elect the participant's own candidate. Our findings show that participants found the MI-agent to be more motivating and empathetic compared to the non-MI agent, leading to more interest in using suggestive feedback for maintaining conversational skills. Our results also reveal the strategies (e.g., building a rubric, negotiation, etc.) participants apply in such collaborative decision-making tasks. Based on the exploration we propose further suggestions that can be used to refine the pre-meeting training or reminder. The contributions of our work are:

- We develop a suggestive motivational interviewing chatbot as pre-meeting feedback for maintaining conversational skills during a remote group discussion. The MI-agent outperforms a non-MI agent in terms of likability, empathy, motivation, etc. Our findings reveal that MI can be a valuable characteristic of feedback agents to improve collaboration.
- We design a résumé-reviewing user study with both team and personal goals to analyze the decision-making strategies participants adopt in such discussions. Our analysis reveals that if conversational skills and decision-making strategies are not applied properly, then even with ideal circumstances one may not succeed in achieving the optimal outcome. We propose that caution and suggestion is needed in pre-meeting training to become more collaborative.

2 RELATED WORK

We explore the literature in terms of which conversational skills are associated with more collaborative behaviors for group discussion, how the videochat systems are being used and improved over time, and what feedback strategies have been applied in developing feedback agents.

2.1 Collaboration Dynamics

During a group discussion, maintaining the appropriate conversational skills can influence the successful exchange of ideas [44]. Features such as balanced speaking [17, 18, 39, 72], less interruptions [20], maintaining mutual respect [42, 43, 62], and collaborative language usage [29, 33] have been emphasized for effective group discussions. Balanced speaking involves taking the speaker's floor when needed as well as releasing the floor to receive information from others in the team. Without active participation, members may not be able to convey important information crucial to the decision-making and may also perceive the meeting to be less effective and collaborative [11]. Releasing the speaker's floor for active listening is important to understand the viewpoints of other members and build upon them. Burgeon and Hoobler [10] observe that over participation or inflated speaking time can be perceived as dominance, leading to reduced team satisfaction. Therefore, it is important to be careful about both over and under participation. Another feature often associated with dominance is intentional impolite interruption. Based on the intent and the execution of an interruption, it can appear as constructive or obstructive [58, 64, 73]. For example, polite interruptions can be successfully applied for *back-channeling* and *handover-request*, especially during remote discussions [68]. Thus the knowledge of how to politely interrupt to take the speaker's floor, as well as what to do upon being interrupted, can be valuable. Group discussion may also involve exchanging conflicting ideas or disagreeing with the other members. If mutual respect is not properly maintained during disagreements, it can misdirect the discussion leading to an ineffective and unsatisfactory meeting [27]. Therefore, how to disagree without disrespecting other teammates is a crucial skill to have. Research has also found that frequent use of first-person pronouns can be perceived as self-focused [75], whereas using more collaborative pronouns (e.g., 'we') can be seen as more team-focused. And finally, positive collaboration is influenced by showing respect towards teammates and keeping up an inclusive environment [16, 42]. Thus helping other members who may be unsupported can keep the team environment safe and positive. Based on the research discussed above, we prepare suggestive feedback for each of these features for users.

2.2 Feedback for Videochats

The proliferation of technical devices and internet access has increased the usage of videoconferencing for personal and professional meetings [9, 25, 71]. However, while discussing over remote mediums, attendees may pay less attention towards social signals [31, 77] and the task itself [13]. Feedback on conversational skill features has been found effective in improving group discussion dynamics [19, 22, 34, 35, 40, 60, 63–65]. Especially for video chats, real-time and post-meeting feedback has been applied for self-reflection [12, 22, 45, 50]. Calacci et al. [12] presents the system

Breakout that consists of real-time visualization of participation feature of each attendee. Samrose et al. [64] shows that real-time feedback on participation or speaking time, vocal volume, interruption count, and facial emotion can be useful during heated discussions over videochat. Faucett et al. [23] introduces *ReflectLive*, which delivers real-time feedback on non-verbal behaviors for clinician-patient videoconferencing. Ali et al. developed *LISSA* that provided real-time feedback to participants to improve their social skills [2]. However, studies have also found real-time feedback to be distracting, potentially increasing the user's cognitive overload and negatively impacting the ongoing discussion [53, 74, 76]. Post-meeting feedback systems have also been explored with a view to improving conversational skill awareness for videoconferencing meetings. *EMODASH* [22], an interactive dashboard providing feedback on affective features from a videochat meeting, was found to successfully improve behavioral awareness over time. *CoCo* [65], an automated post-meeting feedback system providing performance feedback on participation, turn-taking, speech overlap, valence, etc. through a chatbot after a videoconferencing based meeting, was able to bring balanced participation to the team. *MeetingCoach*, which is a post-meeting feedback dashboard, showcased summarized, suggestive, and temporal feedback on meeting dynamics of remote workplace discussions. Their study mentions that users expressed the need for actionable suggestive feedback as reminders before the meeting. Pre-meeting training can be useful to remind the users about the conversational skills and thus improve the collaboration dynamics. Even though not extensively explored in a videochat context, suggestion-oriented feedback has been found effective in behavior modulation [56, 57, 74, 76]. Rhema [74] and ROC Speak [1, 24] systems successfully improve public speaking performance through real-time and post suggestive feedback. Tausczik et al. [76] also provide suggestions during text chat and find that sometimes it can be difficult for users to modify the behaviors on the go. Therefore, research regarding how to effectively provide pre-meeting suggestive feedback to teams remains open for deeper exploration.

2.3 Agent with Motivational Interviewing

Motivational interviewing (MI) [48], a directive counseling method for encouraging commitment towards behavior change, has been effective especially in high-risk scenarios (e.g., moderating alcohol consumption). Low-risk scenarios such as group conversations (e.g., interruption, domination) may still require changes through guided interventions. This can be especially important for agents providing feedback on user behavior or skill, as *self-affirmation* and *defensive bias* theories indicate that feedback may appear confrontational inducing resistance towards behavior change [69, 70]. Related literature on MI-techniques applied in both computerized and non-computerized setups [26, 37, 59] have found MI effective in bringing change to habits, especially through *MI treatment integrity code (MITI)* [46], *stages and techniques of MI* [37], *Manual for MI Skill Code (MISC)* [47], *Drinker's Check-Up (DCU)* [49]. Recent work with MI-based virtual agents [5, 54] use limited response options for users by providing menu-based text-only options. Since MI requires deeper elicitation from the individual needing the assistance, because of the limitation in current agent capabilities in continuing open-ended conversation the automated MI interventions is challenging [45]. Thus, such limited response-based interactions, even though help users, still hold the opportunity to be better in terms of seamless human-machine interaction. We investigate how to make feedback systems and agents more interactive and empathetic so that the information delivered to the users is easily understandable, effective, and motivating.

3 WIREFRAME PROTOTYPE

As the first step towards developing a conversational skill improvement agent using suggestive MI, we (1) construct the preliminary dialogue structure of the agent, and (2) design a paper prototype of the chatbot system.

3.1 Agent Design

Dialogue Construction: To build the dialogue structure of the virtual agent, we worked with the Center for Student Conflict Management at University of Rochester regarding the techniques they follow to provide feedback to people with various forms of misbehavior or misdemeanor. Alongside, we explored the related literature on MI techniques applied in both computerized and non-computerized setups [26, 37, 59]. Highly focusing on *MI treatment integrity code (MITI)* [46], *stages and techniques of MI* [37], *Manual for MI Skill Code (MISC)* [47], we construct the primary language and conversation flow of the agent by following OARS and DOGAN-CAT based change talk. In this preliminary stage, following the computerized MI-based *Drinker's Check-Up (DCU)* [49], our design includes limited response options for users by providing menu-based text-only options. Such limited-option based user response style has been also adopted in recent work with MI-based virtual agents [5, 54].
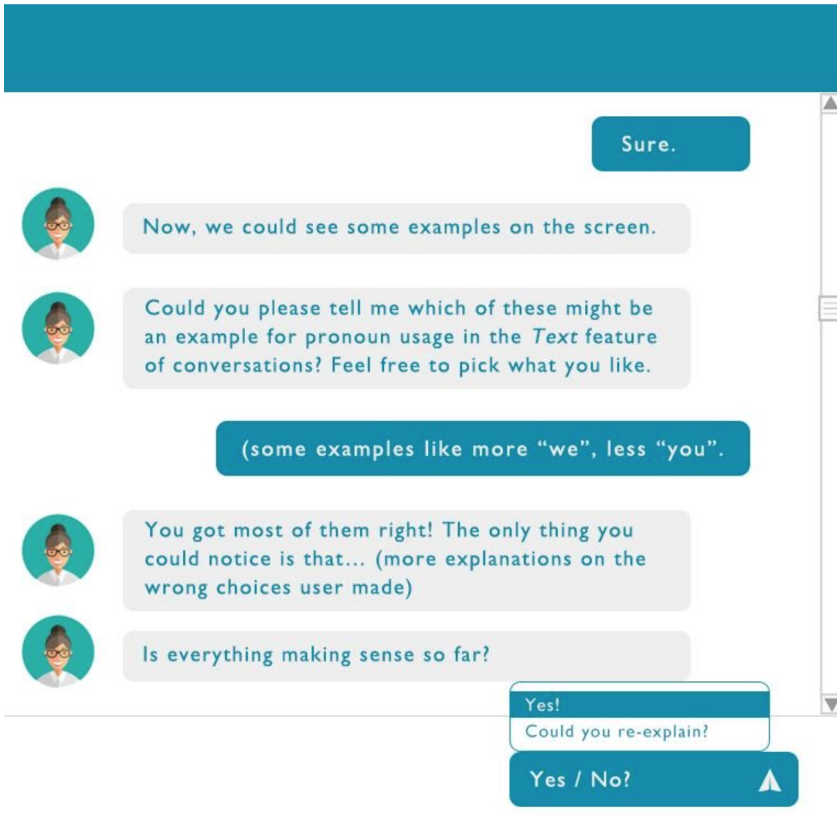

Fig. 1. Prototype of MI-based suggestion incorporated feedback chatbot. In this version, the user responses are menu-based text input.

Chatbot Agent Selection: Recent research has started exploring MI-based virtual agents (e.g., chatbot, embodied agent, social robots) [5, 54, 55] for evoking habit changes. However, in those setups, user interaction is still processed mostly as text-based input, if not through Wizard-of-Oz. The reason that MI-based agents are becoming an emerging research direction is because chatbots reduce the confounding factors beyond MI, such as agent's vocal tone, appearance, speech recognition capabilities. Therefore, to investigate MI through semi-open ended user response for improving conversational skills, we choose a chatbot agent for the interaction. Figure 1 shows the chatbot prototype with agent property and user interaction example.

4 SUGGESTION INCORPORATED MI CHATBOT SYSTEM

We refine the dialogue structure and revise it with the Center for Student Conflict Management in two rounds. Table 1 exemplifies the MI strategies adopted in our refined agent in detail.

Dialogue-based human-agent interaction can depend on not only the dialogue structure but also the dialogue flow. The interaction needs to follow a cohesive and personable narrative keeping the discussion on track. As for the dialogue flow, we incorporate the following strategies in the MI agent:

(1) **Hierarchical Transition Network:** The dialogue is managed by a hierarchical transition network (HTN) with branches in the dialogue based on user response [7]. For example, based on the current commitment level of the user, the agent decides to either move onto the next change-talk phase (if commitment is established) or probe further on the current level (if low commitment).

(2) **Structural Flow:** The dialogue discourse follows 4 stages: (1)introduction, clarity, and rapport building, (2) motivational interviewing, (3) suggestion-based feedback on conversational skills, (4) summary, commitment review, and conclusion.

(3) **Semi-Open Endedness:** Introducing a mix of menu-based text-input (which keeps the discussion on track and controls the flow as intended) and textbox-based open input (which allows users to express their opinion in their own words). For some open textboxes, we invoke guided change talk; for example to induce *change-talk: ability*, the user textbox starts with *"I can __<textbox input>__"*.

(4) **Suggestions:** We build a list of suggestions regarding conversational skills to be delivered. We include specific examples related to 6 categories: *Taking speaker's floor, Releasing the speaker's floor, Disagreement, Interruption, Pronouns usage, Alliance.*

(5) **Clarity of Limitations:** During agent introduction, the dialogue includes transparency regarding agent limitations with sensing and comprehension. For example, *Agent: "My capabilities can be limited sometimes, but I'm here with you to help as much as possible."*

As discussed in the related literature, the features of conversational skills we target are *balanced participation, handling disagreement, collaboration in terms of language, and reduced interruption.* Based on the techniques associated with conversation handover [36, 68] and mutual respect [16, 43], we prepare a set of suggestions as feedback on the features:

(1) **Balanced Speaking:**
   (a) Taking speaker's floor: *"To get the speaker's floor, politely interject with sentences like - 'If I can add something here, ___' "*
   (b) Releasing speaker's floor: *"To release the speaker's floor, one way is to ask an engaging question - 'What do you all think about this idea?' "*

(2) **Handling Disagreement:**

(a) Stating disagreement: *"Respectful disagreement is possible through active listening. Summarize others' points, find common ground, then add your perspective. For example, 'I agree with your point about ___, I would also like to point out that ___.'"*

(b) Facing interruption: *"Upon being interrupted, acknowledge it and clarify your timing needs. For example: 'If I can quickly finish up my thoughts, ___.'"*

(3) **Being Collaborative:**

(a) Pronoun usage: *"Studies show that collaborative pronouns bring a sense of togetherness. For example, using 'We' instead of 'I'."*

(b) Alliance: *"Keeping an eye on whether others are able to express their opinions establishes a safe space. For example, 'Samantha got interrupted before, let's hear her ideas on this'."*

(4) **Summary:**

(a) Motivation summary: *"So, as a final step, let's summarize the plan. You expressed that conversational skill development is important to you."*

(b) Suggestion summary: *"Summary of the tips we discussed are:*
Take speaker's floor -> 'If I can add something here'
Release speaker's floor -> 'What do you all think about this?'
Disagree -> 'I agree with your point-X, but want to add point-Y'.
Pronouns -> More 'We' instead of 'I'.
Offer help -> 'Let attendee-A finish up the thought.' "

Table 1. Example showing MI strategies incorporated in our user-chatbot interaction

Strategy	Example
OARS Technique	
Open Questions: Asking questions for more details to encourage further active involvement in the conversation.	Agent: *"Can you tell me about the last group discussion you were dissatisfied with? What made you feel that way?"* User: *<Input>* (Type-in Option)
Affirmation: Commenting positively on the strengths and efforts.	Agent: *"Learning to be better at group discussions needs effort. You've already taken the first step by being in this session."*
Reflection: Repeating or rephrasing the user response to showcase empathetic listening	Agent: *"So if you decide to improve those skills, how would you do it?"* User: *"I can <input>"* (Type-in Option) Agent: *"So you would <extract/shorten user input> and do much more to make progress. Awesome! I can help with finding tips and tricks to improve conversational skills. Let's dive deeper, shall we?"*
Summarizing: Pulling together core change-talk discussions for consideration.	Agent: *"So, as a final step, let's summarize the plan. You expressed that conversational skill development is important to you. Summary of the tips we discussed are:"* (Summary shown)

Change Talk: DARN CAT[2]	
Ability: Personal perception of ability.	Agent: *"So if you decide to improve those skills, how would you do it?"* User: *"I can <input>"* (Probing for ability change talk)
Reason: Rationale or incentive for making changes.	Agent: *"Can you give a reason why it is important to you?"* User: *"If I improve my conversational skills, then <input>"* *(Probing for reasoning)*
Need: A level of necessity or urgency for making changes; may not have reasoning.	Agent: *"On a scale of 1 (low) to 10 (high), how important is it to you to improve conversational skills in order to help yourself or others in a discussion?*
Commitment: Agreement towards future target changes.	Agent: *"We can set a goal to apply these tips in the next meeting.* *...* User: *"Sounds like a plan to apply!"* (Response Option)
Action: Current movement or confidence towards change.	Agent: *"On a scale of 1 (low) to 10 (high), how confident are you that you can carry out this plan for the next week?* User: *"<Input>"* (1-10 Scale Options) *"You don't have to be 100% confident to get started, you just need enough confidence to take the first step. Give it a go!"*

So far we have discussed the dialogue properties of the MI-agent by showing a discrete example. In Fig 2, we show a snippet of the continuous dialogue of agent-MIA. The snippet exemplifies how commitment is measured and motivation is developed. For example, when the user mentions ways in which s/he plans to bring changes, MIA summarizes the views demonstrating active listening, and then brings up follow-up discussion on other helpful ways for the user to achieve the goal. Based on the commitment level of the user, MIA can decide to assert change-talk differently, as shown in the dialogue snippet in Fig 2.

4.1 Agent Development: MI agent "MIA" VS non-MI agent "Roboto"

To compare our MI agent "MIA" through a user study, we develop a non-MI agent "Roboto" as the control case for our treatment-control study. The dialogue set of "Roboto" is inspired by the interaction style of previous research [65, 76] which design the chatbot dialogue to be engaging as well as informative. As our main confound is MI in agent property, both of these agents contain suggestion based feedback strategies on the same conversational skill categories. However, the non-MI agent does not hold the MI strategies such as asking for a detailed overview from the users in their own words. Notably, the interaction style of the currently available virtual agents is mostly transactional and at times encouraging, but without much focus on evoking motivation through open discussion. Our non-MI agent also shares this style and thus resembles currently available and widely used agent's basic dialogue style.

More specifically, the interaction properties of the non-MI agent "Roboto" are:

---

[2]'D: desire' and 'T: steps takes before' not included in our design

Assess commitment level	**Agent:** On a scale of 1 (low) to 10 (high), how important is it to you to improve conversational skills in order to help yourself or others in a discussion?
Close-ended Likert scale	**User:** 1 | 2 | … | 10 (e.g., user selects number X)

Reflection and continuation based on user's previous response

If X<5: Low Commitment. Evoking stronger change talk by reminding intense cases.	Open-endedness allowing full control	**Agent:** You selected 2. It sounds like improving conversational skills is somewhat important to you. If you *don't* improve anymore, how do you think it will affect your productivity and performance?
		User: ___<text box>___
If X>=5: high commitment. Probing for 'reason' based change talk.	Semi-open-endedness for probing	**Agent:** You selected 8. It sounds like improving conversational skills has value to you. Can you give a reason why it is important to you?
		User: If I improve my conversational skills, then ___<text box>___

Reflection + 'Ability' change talk	**Agent:** Alright, so what I'm hearing you say is that you would like to reduce some concerns. Having the desire to bring changes helps us grow! So if you decide to improve those skills, how would you do it?"
Semi-open-endedness probing for 'ability' based change talk	**User:** I can __<text box>__ (e.g., user types = practice more and more)
Reflection rephrasing user response, showing active listening, and defining next steps	**Agent:** So you would practice and do much more to make progress. Awesome! I can help with finding tips and tricks to improve conversational skills. Let's dive deeper, shall we?"

Fig. 2. A continuous dialogue snippet as an example of agent-MIA invoking motivation through OARS and change talk strategies. Notably, as for the chatbot property, we include open-endedness, close-endedness, and semi-openendedness to achieve specific purposes, as shown in the example. The commitment level of the user is assessed at various interaction points for the agent to strategize better.

- **Informative:** The main purpose of any informative agent is to deliver information in an understandable manner. Based on the understanding level, more information can be provided to improve the user's knowledge base. Examples include information retrieval agents, recommender agents, and so on [30].
- **Transactional:** The interaction need for most currently available agents follows the transactional property. The notion includes providing a service to the user as quickly and accurately as possible without the need for rapport or engagement. Studies have explored such agent dialogues interacting in a *succinct* manner without the agent asking much follow-up or reflection [61]. The response option a user receives is mostly closed-ended, which is an

Dialogue Property	Dialogue Flow Snippet: MIA	Dialogue Property	Dialogue Flow Snippet: Roboto
Assessing level of commitment	**Agent:** We can set a goal to apply these tips in the next meeting. On a scale of 1 (low) to 10 (high), how confident are you that you can carry out this plan for the next week?	Assessing level of commitment	**Agent:** Do all these suggestions seem important to you?
Closed ended (Numeric scale)	**User:** 1 | 2 | … | 10	Close ended (Semantic scale)	**User:** Yes, a lot | Maybe some of them | Not much
MI: Affirmation	**Agent:** You don't have to be 100% confident to get started, you just need enough confidence to take the first step. Give it a go!		_
Closed ended	**User:** Awesome!		_
MI: Summarization (Plan of action + What the user expressed about importance + Suggestion summary)	**Agent:** So, as a final step, let's summarize the plan. You expressed that conversational skill development is important to you. Summary of the tips we discussed are: Take speaker's floor -> 'If I can add something here' Release speaker's floor -> 'What do you all think about this?' Disagree -> 'I agree with your point-X, but want to add point-Y'. Pronouns -> More 'We' instead of 'I'. Offer help -> 'Let attendee-A finish up the thought.'	Empathy + Rapport	**Agent:** I hope you will try to apply some of the suggestions in your future discussions.
Closed ended	**User:** Sounds like a plan to apply!	Close ended	**User:** Sounds like a plan!
Conclusion	**Agent:** Fantastic! Can't wait to hear how the next meetings goes. Good luck, Sonia!	Conclusion	**Agent:** All right then! Good luck with all your discussions, Sonia!
Closed ended	**User:** Thanks, Mia! Bye.	Close ended	**User:** Thanks, Roboto! Bye.

Fig. 3. Comparison of the dialogue of the two agents on the same topic.

important interaction property of transactional agents. An example in our case is: *User: "Show me another example." Roboto:"Another example can be - 'I value your idea on this, if I can share a different perspective ___."' User: "Sounds good!"*

- **Empathetic:** A non-MI agent can still hold empathetic and rapport-building properties [65]. To strengthen the capabilities of "Roboto", we adopt the dialogue style of empathetic agents from previous studies [65, 76]. An example in our sample is: *Roboto: "I hear you. I have some suggestions on being more collaborative."*

In Fig 3, we show a comparative snippet with the dialogue properties of "MIA" and "Roboto". The former helps the user comprehend their own needs and focuses on actions to facilitate changes in

user behavior, whereas the latter focuses on delivering the suggestions in an easily understandable manner without many follow-ups or reflections.

We develop the systems by using HTML, CSS, Javascript, JSON. The system is hosted over a GitHub site and can be accessed from anywhere. Upon loading a system in a browser, the system first asks the username, tracks the user choices and input responses throughout the interaction, and once the conversation ends then downloads the interaction flow file. The username is taken as an input so that both agents can address the user by that name during their interactions. Figure 4 shows the interactive interface of the two agents.

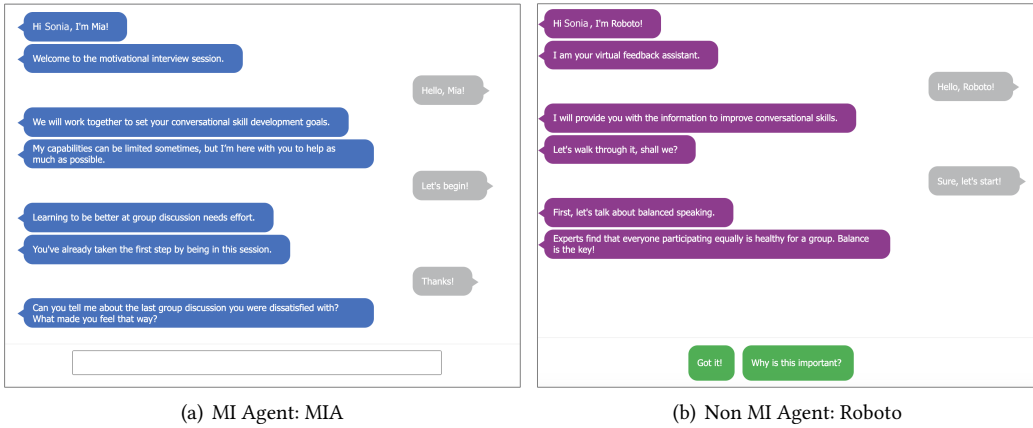| (a) MI Agent: MIA | (b) Non MI Agent: Roboto |

Fig. 4. Chatbot system with and without MI strategies. (Left) The MI-based suggestion incorporated chatbot names "MIA" contains both open-ended (textbox) and close-ended (menu-based) response options for users. (Right) The non-MI suggestion incorporated informative chatbot named "Roboto" includes only menu-based options for user input. Both chatbots hold specific suggestions; however, only agent-Mia adopts motivational interviewing strategies.

5 USER STUDY

5.1 Participants

We recruited 21 participants randomly assigned into 8 teams ($members_{min} = 2, members_{max} = 3$). As per the self-report, the ratio of male and female participants was 10:11. 78.4% reported to be Asian, 14.3% White, 4.8% African American/Africa/Black/Caribbean, 4.8% Hispanic/Latino, and 4.8% Mixed Racial.

5.2 Measures & Study Flow

We conduct a treatment-control-based within-subject counterbalanced experiment. MIA works as the treatment and Roboto as control, and by using a within-subject study we have the same participants evaluating both agents. To remove the impact of the order in which participants interact with the agents, half of the group interact with MIA first and the rest with Roboto first. All participants attended the study remotely over Zoom video call, along with a researcher who also joined the call and managed the study. Throughout the study period, the participants were physically alone in their rooms. During the group discussion task, the audio-video feed of the researcher remained disabled. In the study, each participant interacted with an agent (MI or non-MI agents) and then participated in a group discussion. After that, they interacted with the other agent

followed by another discussion on a similar topic. During every interval (after agent interaction and after discussion), a survey followed measuring their motivation, satisfaction, perception, etc. We use validated questionnaires from the Assessment of Empathic Communication in medical interviews (REM) [52], and Intrinsic Motivation Inventory Scale (IMI) [15], MI-agent evaluation questionnaires by Olafsson et al. [54]. In our surveys, we also added questions asking the participants to rate and compare the agents as well as comment on the discussion sessions.
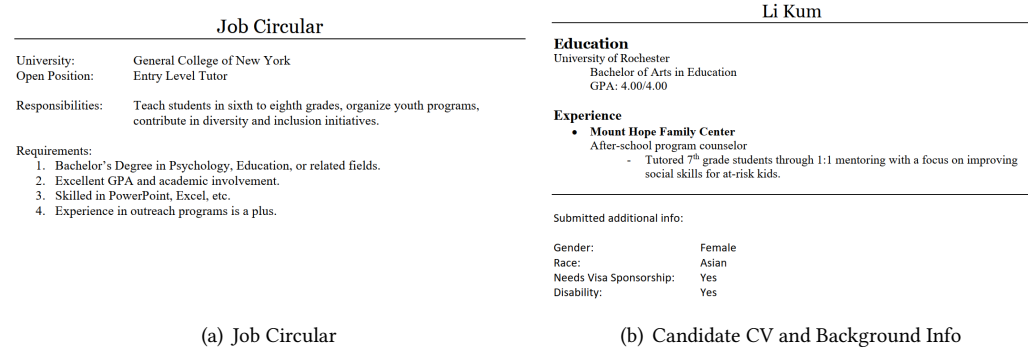
<table>
<tr><td>

Job Circular

University: General College of New York
Open Position: Entry Level Tutor

Responsibilities: Teach students in sixth to eighth grades, organize youth programs, contribute in diversity and inclusion initiatives.

Requirements:
1. Bachelor's Degree in Psychology, Education, or related fields.
2. Excellent GPA and academic involvement.
3. Skilled in PowerPoint, Excel, etc.
4. Experience in outreach programs is a plus.

</td><td>

Li Kum

Education
University of Rochester
   Bachelor of Arts in Education
   GPA: 4.00/4.00

Experience
- **Mount Hope Family Center**
  After-school program counselor
  - Tutored 7th grade students through 1:1 mentoring with a focus on improving social skills for at-risk kids.

Submitted additional info:

Gender: Female
Race: Asian
Needs Visa Sponsorship: Yes
Disability: Yes

</td></tr>
<tr><td>(a) Job Circular</td><td>(b) Candidate CV and Background Info</td></tr>
</table>

Fig. 5. Discussion topic of the teams: sample job circular and candidate résumé that best fits this circular

5.3 Group Discussion Topic

With help from the Greene Center for Career Education & Connections at University of Rochester, we prepare a recruitment scenario for the group discussion topic. For each discussion session, we designed a job circular notice and candidate résumés with some background information (gender, race, disability, visa sponsorship requirement) which the candidates presumably submitted through a cover letter. The Career Center ranked the candidate résumés matching the job circular and also explained their reasoning and suggestions behind the ranking. Their suggestion emphasized that candidates' qualifications and experiences should be considered the most for any recruitment. The ranking was based on how the information in the résumés best fit the requirements mentioned in the job circular. The task of the group discussion was to unanimously select one candidate to offer the job. Each participant was randomly assigned a résumé. Besides that team goal, participants also had the personal goal to advocate for their assigned candidates and convince the team members in electing that candidate. The participant whose candidate gets unanimously elected by the team would receive 10 points and the other members would receive 5, however no consensus would lead to 0 points for all team members.

6 RESULTS

6.1 Agent Evaluations

In the final survey, participants compare the overall qualities of the agents. The survey collects user-preference towards MIA, Roboto, both agents, or none of the agents. In this stage, participants reported preferably towards agent-MI in terms of building up motivation and confidence. Figure 6 shows that while comparing among 'Mia', 'Roboto', 'both', and 'none' conditions, participants liked MIA more and wanted to use it in the future (13 participants out of 21).

Even though both agents have similar dialogue styles except for the MI-interaction component, empathy level is rated highly in favor of agent-MIA, revealing that MI would bring a more empathetic

component to the agent's characteristics. Notably, the knowledge levels of both agents were rated almost equally (8:Mia, 6:Roboto, 7:Both, 0:None; out of 21), which captures our design goal that the knowledge level and delivered feedback information by both agents were almost the same; and the only confound we have is the motivational interviewing technique and high level of empathy incorporated in the dialogue of agent-MIA. We would like to note that, another variable is the length of the agent interaction, as interaction with agent-MIA takes 7-12mins in comparison with agent-Roboto taking 5-7mins. In our design, we keep this difference, as agent-Roboto should be closer to the traditional virtual agents which provide relevant information in an understandable but concise manner.

These findings above answer which agent is preferred (e.g., "which agent did you like more?", implying if any). As we establish that one agent is indeed being preferred by participants (instead of 'both' or 'none'), we now measure how much the preference leans towards each component. After each interaction with the agent, participants filled out a survey rating each individual agent on a 7-scale Likert chart, which we compare across agents to reveal whether the rating differences are significant or not. We conduct a repeated measure ANOVA on each rating per category (helpfulness, likeability, etc.). It shows statistically significant results for agent-MIA being more motivating ($mean_{mia} = 5.1, mean_{roboto} = 4.14, F(1, 40) = 9.78, p < 0.05$), satisfactory ($mean_{mia} = 5.05, mean_{roboto} = 4.09, F(1, 40) = 12.31, p < 0.05$), likeable ($mean_{mia} = 5.0, mean_{roboto} = 4.24, F(1, 40) = 5.32, p < 0.05$). For other features, such as helpfulness Mia was found to be more helpful than Roboto (Mia: 48%, Roboto: 23%, Both: 29%), however the difference was not statistically significant. Notably, knowledge level and coherence differences were not statistically significant, which confirms our design principle of making both agents similar in terms of those characteristics.
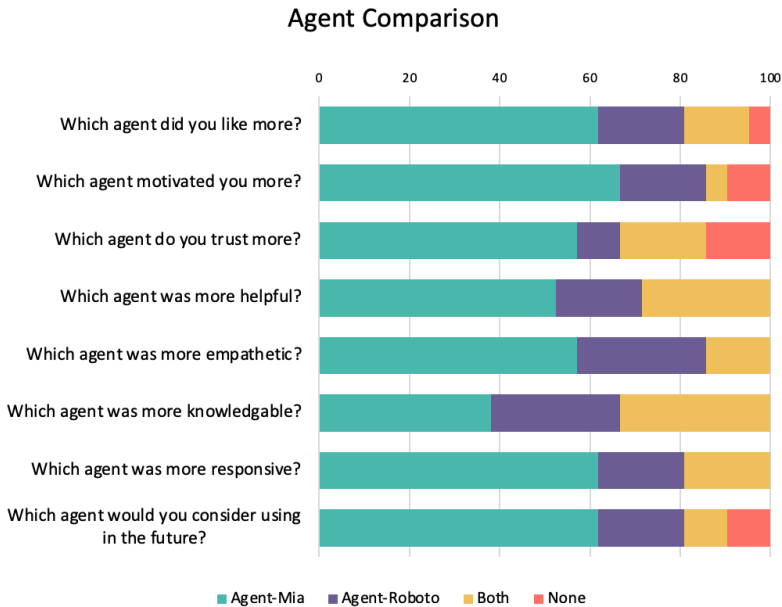

Fig. 6. Comparison of efficacy and trust of the MI and the non-MI based agents

6.2 Suggestion/Feedback Usage

Participants evaluated the usefulness of the example-based suggestions they received on 6 conversational skills (how to: *take speaker's floor, release speaker's floor, handle being interrupted, disagree, use the collaborative pronoun, be an ally*. Figure 7 shows the ratings on a 7-scale Likert (1-7: 1-not at all, 7-a lot), which shows that all categories were rated useful, with 'handling disagreement' receiving the highest average score of 5.48. The participants also reported which suggestions they applied more and which ones they hoped their teammates could have applied more. Notably, each of the responses can mention multiple features, so the percentage of people bringing up each feature would not necessarily add up to 100%. 57% of participants reported that they applied the suggestion on how to handle disagreement, 36% how to take speaker's floor, 29% how to release the floor. On the other hand, 26% expressed that their team members maintained their conversational skills well. 24% hoped that the teammates applied the suggestion of handling disagreement even more.

6.3 Analyzing Discussion Outcome

We analyze the discussions performed by the participants to capture the attributes contributing to electing the final candidate in the task. Out of the 16 discussions, 7 successfully elected the *actual* best candidate as their final choice (44%), 4 elected the *actual* second-best candidate (25%), 4 ended up electing the *actual* worst candidate (25%), and 1 discussion decided not to elect anyone (6%). Through a qualitative lens, we dissect how the decisions were being made.

Notably, having the best candidate profile does not guarantee that a participant would be able to convince the teammates to select that candidate. This is because even though this type of group discussion resembles debates, there are other key conditions that influence the consensus. For example, in debate, the decision is highly and mostly influenced by the reasoning presented, whereas in public political debate, the style of reasoning is seen as important in reaching the target audience[3]. We show that for groups, rapport or conflict among other members can also influence someone's own performance. We applied Grounded Theory to utilize the transcript of the group

---

[3]https://www.nytimes.com/2020/10/06/us/likability-ambition-kamala-harris-debate-mike-pence.html
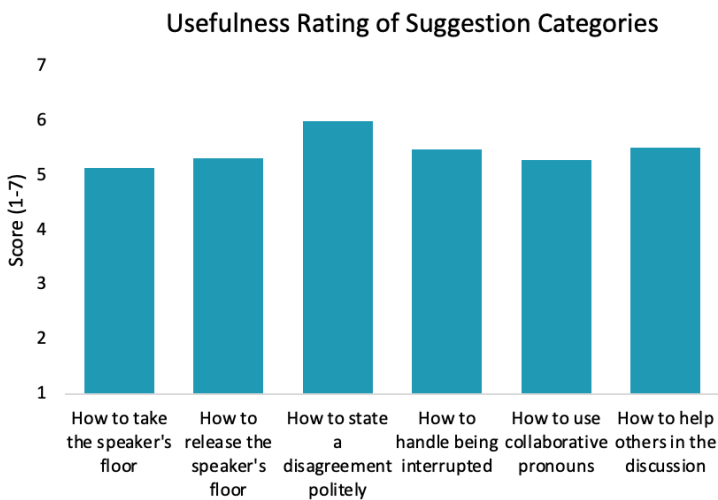

Fig. 7. Usefulness ratings of the suggestion categories

discussion revealing the potential themes related to collaboration emerging in the discussions and present the exemplars [6]. By using grounded theory, we show that our candidate election group discussion held 4 attributes influencing the team consensus:

- **Objectivity - Argument Strategy:** Just like any debate, argument quality and skill are the major keys to influencing a consensus [38]. With the best candidate, it may be easier to form better arguments, however we found that strategies like controlling the importance of the job requirements, advocating for candidates early, forming alliances, etc. were effective in making a candidate appear more lucrative.

- **Subjectivity - Likability:** Likability is highly correlated with electability [8, 79]. In a team, it consists of 'what' is being said (*collaborative or personable linguistic style*, besides argument quality [51]) and 'how' (*para-linguistic cues* [66]). Maintaining communicational skills is important to ensure likability (e.g., apologizing upon interrupting) and argument effectiveness (e.g., not shying away from taking the speaker's floor to advocate for their cause).

- **Subjectivity - Others' Rapport/Conflict:** Even though self-performance is important, how the other members perform has an impact. If the other members form an alliance through rapport building, they can advocate for a combined cause. If the other members have a conflict with each other, it can benefit the non-conflicting members in electing their candidate or can also cause a disservice to everyone by not reaching a consensus. Therefore, even after maintaining one's own communicational skills, if other members are not maintaining theirs then it might impact both the group and the personal goal outcomes.

- **Other Factors:** The participants were trying to maximize their personal goals (10 for convincing others, 5 for reaching unanimous consensus, 0 for no consensus). Even when they did not fully agree on the final choice, they converged to avoid receiving a zero. Since each discussion time limit was 15 minutes, advocating for their candidates late left little time to convince others.

6.4 Analyzing Decision-making Strategies

In this stage, by using thematic analysis, we categorize the strategies the participants used to reach a consensus. We found 4 strategies: (1) building the rubric: discussing the target requirements and assigning varying weights to each of them; (2) choosing the second-best: everyone's top choice being intact, voting for the second-best candidate; (3) negotiation on personal goals: focusing on how to maximize getting most points from each discussion sessions; (4) other: the case when no unanimous consensus achieved.

- **Building the Rubric:** 10 out of 16 times the teams tried to form a rubric based on the job circular and assigned points for each candidate. The Career Center also suggested this approach for assessing how well the candidates fit the job requirements for finding out the *best/global fit*.
  As per the task, the participants advocated for their own candidates, focusing more on factors in the rubric that supported their own candidates or hurt the opponents'. Controlling the weight assignment in a way that highlighted the strengths and downplayed the weaknesses of a candidate could create a powerful narrative. However, reaching a consensus based on fulfilling the rubric requirement appears logical; during surveys for these discussions no member reported dissatisfaction towards the group decision. A snippet from such a discussion of group-1, in which P1 had the *actual* second-best, P2 had *actual* best, P3 had *actual* third-best candidates:

P1: *"The job circular needs someone who has experience in STATA, I think my candidate has the better experience in it."* [Note: assigns high weight]

P3: *"Can we go towards some objective measures like industry experience, CGPA, statistics? We can come to an agreement about which one should get the job."* [Note: Rubric proposed]

P1: *"That's a good point."*

P3: *"My candidate is okay for the responsibility. He has skills in statistics but doesn't have the experience. He has the best CGPA (4.0). So I will give him a score of four out of five."* [Note: interest in global best]

P1: *"My candidate is skilled in statistics, has experience in working in the industry, has leadership skills, and has a good CGPA (3.80)"* [Note: unlike P3, highlights strength]

P3: *"If I can add, 3.80 is also sufficient CGPA."*

P1: *"Yes, yes!"*

P3: *"I am convinced that your candidate has all the criteria fulfilled."*

P2: *"Sorry for interrupting, my candidate has experience as a teaching assistant in his statistics course. So I think it's fair to assume that he has a deep understanding."* [Note: waited too long to interject]

P1: *"But there is no mention of STATA in his CV."* [Note: expresses concern]

P2: *"My candidate has internship experience that directly matches the job description. So I feel like he would literally be transitioning from the internship to a full-time job."* [Note: high-weighted category, but concern unaddressed]

P1: *"I'm not convinced; he has a little lacking on that part (STATA)."*

P3: *"I think P1's candidate is the best."*

The above interaction reveals that if a concern is expressed, it is important to address it before bringing up new points, which requires active listening. Also, rather than providing all the info at once, it is important to participate in each stage of a discussion (weight assignment, voting, etc.). Here is another example from group-2 that successfully picks the *actual* best candidate:

P4: *"... They (candidate-1) have the industry experience, they have the math experience with market stuff. I'm not a financial person but (laughs), yeah I think they are really a strong candidate."*

P5: *"Ok guys, I'm gonna be real honest here, I think candidate-1 is the best option (smiles). They have, as she said before, the skills... they don't even need to learn the job requirements."*

P6: *"It's true, but not to discount the importance of being familiar with STATA. I don't know what it is, but I think it is a complicated piece of software, it probably will take a while to be brought up to speed and get used to it. John (candidate-3) has a perfect GPA and experience with the tools, so I think he is a really good choice for hitting the ground running"*

P5: *"I just looked into it and STATA is actually a software people use for statistics and data science."*

P6: *"Yeah, I can imagine it takes some time to learn."*

P4: *"I'm sure there is a learning curve with that, but I feel like given that my candidate had majored in this field I'm sure they have come across that software. I'm not like positive but there is a good chance, I mean in the same way that you just come across R in general stat class here... Plus, the circular doesn't say we have to have that one particular software (STATA), it's more like those are kind of pluses as long as you have the general skills, and it seems that my candidate can do those general skills."*

P6: *"Yeah, I agree then too that specific software is not super important. I think candidate-1 does seem like the best candidate here."*

This interaction above reflected active listening as it successfully addressed the concern raised by others and stayed on the topic before the concern was resolved. Alongside, the discussion readjusted the weights of the metric to reflect global fit instead of benefiting local best options. The participants took the speaker's floor and released it when necessary and effectively handled disagreements through active listening.

Therefore, we propose that to find the optimal/global decision through a group discussion – *(1) Care should be taken towards how the requirements (rubric and weights) are developed and evaluated, (2) Active listening should be used while expressing disagreements.*

- **Choosing Second Best:** In 3 cases, participants decided to discuss and vote for their second choice after everyone refused to shift from their first. Convincing at least one other person boosted the prospects of that candidate being elected, as the convinced member also started advocating in parallel and voted that candidate as their second-best choice.

    P8: *"I would not want to exclude candidates who are highly skilled but needs visa sponsorship, just because that anyone should be able to get a job, period (laughs)."* [Note: P8's candidate needs visa sponsorship]
    P9: *"I think we should leave that for the legal team (laughs)."*
    P8: *"Yes (laughs)."*
    P7: *"I think that diversity is also important, oh although my applicant is a white person (laughs)."* [Note: candidate weakness expressed]
    P8: *"If diversity is what you are after, then bringing more females and people of color in the field of economics is the key. So that's why Li Kum (candidate-2) is your girl (laughs)."* [Note: assigns more weight]
    P7: *"I completely agree. Which is why Danzel (candidate-1), a non-binary black person should be our pick... And honestly Danzel proved themselves in work and university. I feel like if we hire them, it's not because of diversity, frankly they are the best among the three."* [Note: mentions global fit]
    P8: *"Okay, I think if Li Kum is not on anyone's radar, then Danzel would be my second choice."* [Note: convinced]

Notably, sometimes participants voted for the least favorite candidate as their second choice hoping the discussion would continue, which backfired. Also, as teams did not want to leave without a unanimous consensus, which would mean zero points for all members, they often voted for the candidate other participants already sided with.

Therefore, we find that to promote one's own agenda during a group discussion – *(1) Listening to what qualities others value and convincing a sub-group can increase the prospects, (2) When the voting or decision stage begins, care should be taken about the discussion requirements (unanimous decision, timing, etc.).*

- **Negotiation on Personal Goals:** In 2 discussion cases with 1 group with 2 participants, after advocating for their candidates but being unable to fully convince the other teammates they negotiated on optimizing their personal score. The motivation to optimize personal goals is the major motivation in other groups as well. Participants shot for 10 points by convincing the group, but they would go for an agreement electing a different candidate and leave with 5 points rather than not reaching a consensus resulting in 0 points.

    P15: *"The job description says 6th to 8th grade, and your candidate says active learning but no specifics on who they are teaching, versus mine who teaches 7th grade which is the right*

> *age group."* [Note: *Has actual best candidate*]
> P16: *"Ya that's something for sure. Both are strong in skill sets."*
> P15: *"We have one minute... I think my candidate is a little bit better suited but I would rather pick someone than not pick anyone (smiles)."*
> P16: *"I think we can pick my candidate. I don't know what the score means... but maybe we can share (laughs)."*
> P15: *"Okay we can pick your candidate then."* [Note: *They pick P15's candidate in next discussion*]

Negotiation on the personal goals may benefit the members but not find the best discussion outcome. During the survey, in the first discussion P15 reported not being quite satisfied with the group decision but found it productive, whereas in the second discussion both reported being satisfied.

Therefore, we point out that for optimizing personal goals during a group discussion – *negotiation can find an outcome satisfying the members' goals, even though that may not bring the optimal discussion outcome.*

- **Other:** In 1 discussion, the group did not reach a unanimous decision, and thus every member received 0 points. The participant who did not converge mentioned not being convinced about the outcome. We mentioned above that maximizing personal goals was the intrinsic motivation in the experiment, so we discuss why this case occurred. The reasons can be an unconvincing argument, previous discussion impact (the member not converging had the actual best candidate in the first session), less rapport with the team members, etc. can contribute to such outcome. During the survey for this discussion, the member not converging reported being satisfied with the group decision whereas the other two were dissatisfied.

  Therefore, we hypothesize that – *maintaining rapport with the team as well as sometimes resolving conflicts of other members are ultimately important for groups with recurring meetings.*

These findings are important as they can be applied to refine the suggestion incorporated feedback of the agent even more, so that attendees in similar decision-making meetings can be better prepared.

7 DISCUSSION

Our study shows that agents with motivational interviewing appeared more empathetic and appealing for future use. Participants perceived to have used the suggestions by agent-MIA more. Notably, agent-MIA still bears limited capacity in terms of reflection. The follow-up reflection it provides to users is of the *simple reflection* category, whereas highly skilled therapists use the *complex reflection* MI strategy. Even though participants appreciated being heard by the agent, they did notice those simplistic responses and expressed that they would like the agent to go deeper. The natural language understanding and generation community can work on incorporating *complex reflection* in the dialogues of virtual agents by better summarizing user's long responses.

> P4: *"I liked that they let me type out answers and asked follow-up questions. I think it's very cool it sometimes "understood" what I was typing (namely when I mentioned practice). I appreciate that they let me chose discussion topics. I also felt they had a friendlier, less preachy tone."*

> P11: *"I liked that Agent-Mia was more responsive and seemed to be more interested in my personal struggles. This made me feel that her advice was better suited for my needs and was almost tailored to me based on my opinions."*
>
> P19: *"The agent did not seem to be understanding the content of my conversation, although it probably understood the sentiment of it. Especially, I did not feel like the agent understood the long responses at all."*

There are a couple of experimental limitations in our discussion study design. The motivation to maximize the point by convincing others about their own candidate was 'soft', as it did not translate into actual gain. For example, providing the winner with additional gift cards could have established stronger motivation to push for their candidates or not select anyone at all. The 15-minute discussion time limit was also a factor restricting the ability to have a deeper conversation. While reviewing the job circular and the resumes, the Career Center mentioned that a candidate's demographic background (e.g., race, gender, etc.) could not be used for the sole proposes of hiring.

Another limitation of our study is the framing of the agents. Agent-Roboto introduces itself as a 'virtual assistant', whereas MIA does not. Framing is important for establishing the social presence and user expectation level of chatbots [4]. We avoid introducing MIA as a 'virtual assistant' (1) so that users do not expect it to interact as general virtual assistants which do not have motivational interviewing capabilities, and (2) as motivation interviewer is expected to act as a 'facilitator or counselor' rather than an 'assistant'. Notably, as the capabilities of the agent are limited and the purpose is not counseling but rather motivation, we also refrain from introducing agent-MIA as a 'counselor'. On the other hand, we introduce agent-Roboto as a 'virtual assistant' as it follows the structure or style of a general informative virtual assistant adopted in other studies [61, 65]. Based on these reasons, we cautiously frame the two agents differently. However, how our applied framing might have impacted the user interaction could have been investigated more. What the exact framing of a motivational interviewing agent should be is itself an important research question that remains to be explored.

In the study, in several cases, the participants used demographic info to promote candidates. Interestingly, some were upset with their own proposed logic (e.g., *"I feel sorry for Li Kum (candidate-2), I did not select her because she needed visa sponsorship and now she might not have a job"*), and some used it to promote their candidate better (e.g., *"I do think that you need some sort of life experience to have the level of empathy required for contributing to diversity and inclusion [which the job asks for]. Speaking as a privileged white male, I think the image of a white male [candidate-3 John Smith whom the other participants were siding with] contributing to diversity and inclusion will have (more impact) than non-binary black man (candidate-1 Danzel Kofi who was this participant's candidate) who would be adding so much more to the table."*).

The discussions were impacted not only by the arguments presented but also by how the participants maintained their communication skills. For example, if they shied away from taking the speaker's floor, they had little time left to persuade. That is why taking and releasing the speaker's floor is important, as well as other communicational skills.

Even though research has found that pre-meeting training and reminders might be useful in improving remote group discussion dynamics [63], extensive exploration has not been done to verify the actual effect. Moreover, pre-meeting training can be important in discussion tasks similar to our recruitment study, as reviewing the decision-making strategies can reveal the further training components needed for the members. For example, members reviewing the résumés may have various biases which can be revealed by analyzing the decision-making strategy just like performed in this study. After that, the required changes can be brought by refining the suggestions and feedback delivering right before a meeting.

Our findings reveal that MI strategies can be a valuable characteristic of feedback agents to improve collaboration. It also reveals that if conversational skills and decision-making strategies are not applied properly, then even with the biggest advantage one may still not succeed in achieving the collaboration outcome.

Another important aspect to consider is the usage frequency of such MI-agents. The MI component is helpful when users do not bear much motivation to modify their behaviors. Once the motivation is established, instead of constant usage, periodic follow-ups can be more helpful. In the future, a longitudinal study running with varying usage frequencies can address such research questions.

As our study finds that in a pre-meeting stage MI strategies within an agent can successfully motivate users to improve their conversational skills, an interesting next step can be exploring the real-time impact of it. Notably, MI strategies require in-depth probing and reflection, which might be highly distracting during an ongoing meeting. Whether an MI-agent can be applied during an ongoing group discussion in a non-distracting manner, and how that might impact other group members and/or the discussion as a whole are interesting research questions open for explorations.

8 CONCLUSION

With the help of related literature and the Center for Student Conflict Management, we design dialogue properties of an MI-agent as well as a non-MI agent - both incorporated with specific example-based suggestions. Our findings show that MI-agent appeared more likable, motivating, and responsive compared to the non-MI agent. We incorporate semi-open-endedness in MI-agent so that users can input their own opinion in their own words. We also include a simple reflection technique to follow up on user input. Our results show that even though users felt heard by the MI-agent, the need for complex reflection was prominent. With the help of the Greene Center for Career Education & Connections, we developed the job-candidate experiment. Our discussion brings the hypothesis that for consensus-based discussion among group members, the outcome depends on argument quality or strategy (similar to debates), likability through conversational skills (similar to political debate), and team rapport/conflict (similar to negotiation). We show the strategies used in reaching a decision and present suggestions regarding maximizing group and personal gain outcomes. Our analysis reveals that if conversational skills and decision-making strategies are not applied properly, then even after having the best option one may not succeed in achieving the optimal outcome. We present the considerations that need to be attended during decision-making group discussions. similar decision-making group discussions. Our findings can improve the way feedback agents approach improving collaboration for remote teams.

REFERENCES

[1] [n.d.]. ([n. d.]).

[2] Mohammad Rafayet Ali, Dev Crasta, Li Jin, Agustin Baretto, Joshua Pachter, Ronald D Rogge, and Mohammed Ehsan Hoque. 2015. LISSA—live interactive social skill assistance. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 173–179.

[3] Keith Anderson, Elisabeth André, Tobias Baur, Sara Bernardini, Mathieu Chollet, Evi Chryssafidou, Ionut Damian, Cathy Ennis, Arjan Egges, Patrick Gebhard, et al. 2013. The TARDIS framework: intelligent virtual agents for social coaching in job interviews. In *International Conference on Advances in Computer Entertainment Technology*. Springer, 476–491.

[4] Theo Araujo. 2018. Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Computers in Human Behavior* 85 (2018), 183–189.

[5] Tessa Beinema, Harm op den Akker, Lex van Velsen, and Hermie Hermens. 2021. Tailoring coaching strategies to users' motivation in a multi-agent health coaching application. *Computers in Human Behavior* (2021), 106787.

[6] H Russell Bernard and Harvey Russell Bernard. 2013. *Social research methods: Qualitative and quantitative approaches*. Sage.

[7] Timothy Bickmore and Toni Giorgino. 2006. Health dialog systems for patients and consumers. *Journal of biomedical informatics* 39, 5 (2006), 556–571.

[8] Samuel D Bradley, James R Angelini, and Sungkyoung Lee. 2007. Psychophysiological and memory effects of negative political ads: Aversive, arousing, and well remembered. *Journal of Advertising* 36, 4 (2007), 115–127.

[9] Jed R Brubaker, Gina Venolia, and John C Tang. 2012. Focusing on shared experiences: moving beyond the camera in video communication. In *Proceedings of the Designing Interactive Systems Conference*. 96–105.

[10] JK Burgeon and G Hoobler. 2002. Nonverbal signals. Handbook of interpersonal communication 3rd. Edition.

[11] Judee K Burgoon and Deborah A. Newton. 1991. Applying a social meaning model to relational message interpretations of conversational involvement: Comparing observer and participant perspectives. *The Southern Communication Journal* 56, 2 (1991), 96–113. https://doi.org/10.1080/10417949109372822

[12] Dan Calacci, Oren Lederman, David Shrier, and Alex 'Sandy' Pentland. 2016. Breakout: An open measurement and intervention tool for distributed peer learning groups. *arXiv preprint arXiv:1607.01443* (2016).

[13] Hancheng Cao, Chia-Jung Lee, Shamsi Iqbal, Mary Czerwinski, Priscilla NY Wong, Sean Rintel, Brent Hecht, Jaime Teevan, and Longqi Yang. 2021. Large Scale Analysis of Multitasking Behavior During Remote Meetings. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–13.

[14] Michelene TH Chi. 1996. Constructing self-explanations and scaffolded explanations in tutoring. *Applied Cognitive Psychology* 10, 7 (1996), 33–49.

[15] Jimmy Choi, Tamiko Mogami, and Alice Medalia. 2010. Intrinsic motivation inventory: an adapted measure for schizophrenia research. *Schizophrenia bulletin* 36, 5 (2010), 966–976.

[16] Jonathan Culpeper, Derek Bousfield, and Anne Wichmann. 2003. Impoliteness revisited: with special reference to dynamic and prosodic aspects. *Journal of pragmatics* 35, 10-11 (2003), 1545–1579.

[17] Esther Derby, Diana Larsen, and Ken Schwaber. 2006. *Agile retrospectives: Making good teams great*. Pragmatic Bookshelf.

[18] E. Ilana Diamant, Brian Y. Lim, Andy Echenique, Gilly Leshed, and Susan R. Fussell. 2009. Supporting Intercultural Collaboration with Dynamic Feedback Systems: Preliminary Evidence from a Creative Design Task. In *Extended Abstracts on Human Factors in Computing Systems (CHI EA '09)*.

[19] Joan Morris DiMicco, Katherine J Hollenbach, Anna Pandolfo, and Walter Bender. 2007. The impact of increased awareness while face-to-face. *Human–Computer Interaction* 22, 1-2 (2007), 47–96.

[20] Norah Dunbar, Amy M. Bippus, and Stacy Young. 2008. Interpersonal Dominance in Relational Conflict: A View from Dyadic Power Theory. *Interpersona : An International Journal on Personal Relationships* 2 (10 2008). https://doi.org/10.5964/ijpr.v2i1.16

[21] Norah E Dunbar and Judee K Burgoon. 2005. Perceptions of power and interactional dominance in interpersonal relationships. *Journal of Social and Personal Relationships* 22, 2 (2005), 207–233.

[22] Mohamed Ez-Zaouia, Aurélien Tabard, and Elise Lavoué. 2020. Emodash: A dashboard supporting retrospective awareness of emotions in online learning. *International Journal of Human-Computer Studies* 139 (2020), 102411.

[23] Heather A. Faucett, Matthew L. Lee, and Scott Carter. 2017. I Should Listen More: Real-time Sensing and Feedback of Non-Verbal Communication in Video Telehealth. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW (Dec. 2017), 44:1–44:19.

[24] Michelle Fung, Yina Jin, RuJie Zhao, and Mohammed (Ehsan) Hoque. 2015. ROC Speak: Semi-Automated Personalized Feedback on Nonverbal Behavior from Recorded Videos. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Osaka, Japan) *(UbiComp '15)*. Association for Computing Machinery, New York, NY, USA, 1167–1178. https://doi.org/10.1145/2750858.2804265

[25] Zixiu Guo, John D'ambra, Tim Turner, and Huiying Zhang. 2009. Improving the effectiveness of virtual teams: A comparison of video-conferencing and face-to-face communication in China. *IEEE Transactions on Professional Communication* 52, 1 (2009), 1–16.

[26] Jennifer Hettema, Julie Steele, and William R Miller. 2005. Motivational interviewing. *Annu. Rev. Clin. Psychol.* 1 (2005), 91–111.

[27] Pamela J Hinds and Mark Mortensen. 2005. Understanding conflict in geographically distributed teams: The moderating effects of shared identity, shared context, and spontaneous communication. *Organization science* 16, 3 (2005), 290–307.

[28] Mohammed Hoque, Matthieu Courgeon, Jean-Claude Martin, Bilge Mutlu, and Rosalind W Picard. 2013. Mach: My automated conversation coach. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*. 697–706.

[29] David A. Huffaker, Roderick Swaab, and Daniel Diermeier. 2011. The Language of Coalition Formation in Online Multiparty Negotiations. *Journal of Language and Social Psychology* 30, 1 (2011).

[30] Carlos Hurtado, Margarita Ramirez Ramirez, Arnulfo Alanis, Sergio Octavio Vazquez, Beatriz Ramirez, and Esperanza Manrique. 2018. Towards a multi-agent system for an informative healthcare mobile application. In *KES international symposium on agent and multi-agent systems: technologies and applications*. Springer, 215–219.

[31] Ellen A Isaacs and John C Tang. 1994. What video can and cannot do for collaboration: a case study. *Multimedia systems* 2, 2 (1994), 63–73.

[32] Matthew K. Miller, Martin Johannes Dechant, and Regan L. Mandryk. 2021. Meeting You, Seeing Me: The Role of Social Anxiety, Visual Feedback, and Interface Layout in a Get-to-Know-You Task via Video Chat.. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.

[33] Hayato Kihara, Shogo Fukushima, and Takeshi Naemura. 2016. Analysis of Human Nodding Behavior During Group Work for Designing Nodding Robots. In *Proceedings of the 19th International Conference on Supporting Group Work (GROUP '16)*.

[34] Taemie Kim, Agnes Chang, Lindsey Holland, and Alex Sandy Pentland. 2008. Meeting Mediator: Enhancing Group Collaborationusing Sociometric Feedback. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work* (San Diego, CA, USA) *(CSCW '08)*.

[35] Taemie Kim, Pamela Hinds, and Alex Pentland. 2012. Awareness As an Antidote to Distance: Making Distributed Groups Cooperative and Consistent. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (CSCW '12)*. 1237–1246.

[36] David S. Kirk, Abigail Sellen, and Xiang Cao. 2010. Home video communication: mediating 'closeness'. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work (CSCW '10)*. 135–144.

[37] Florian E Klonek, Vicenç Quera, and Simone Kauffeld. 2015. Coding interactions in Motivational Interviewing with computer-software: What are the advantages for process researchers? *Computers in Human Behavior* 44 (2015), 284–292.

[38] Deanna Kuhn and Wadiya Udell. 2003. The development of argument skills. *Child development* 74, 5 (2003), 1245–1260.

[39] Harold J. Leavitt. 1951. Some Effects of Certain Communication Patterns on Group Performance. *Journal of Abnormal Psychology* (1951).

[40] Gilly Leshed, Diego Perez, Jeffrey T Hancock, Dan Cosley, Jeremy Birnholtz, Soyoung Lee, Poppy L McLeod, and Geri Gay. 2009. Visualizing real-time language-based feedback on teamwork behavior in computer-mediated groups. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 537–546.

[41] Ian Li, Anind Dey, and Jodi Forlizzi. 2010. A stage-based model of personal informatics systems. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 557–566.

[42] Rousiley CM Maia and Thaiane AS Rezende. 2016. Respect and disrespect in deliberation across the networked media environment: examining multiple paths of political talk. *Journal of Computer-Mediated Communication* 21, 2 (2016), 121–139.

[43] Jane Mansbridge, James Bohman, Simone Chambers, David Estlund, Andreas Follesdal, Archon Fung, Cristina Lafont, Bernard Manin, and Jose luis Marti. 2009. The Place of Self-Interest and the Role of Power in Deliberative Democracy. *Journal of Political Philosophy* 18, 1 (2009).

[44] Poppy L. McLeod, Jeffrey K. Liker, and Sharon A. Lobel. 1992. Process Feedback in Task Groups: An Application of Goal Setting. *The Journal of Applied Behavioral Science* 28, 1 (1992).

[45] Matthew K Miller, Regan L Mandryk, Max V Birk, Ansgar E Depping, and Tushita Patel. 2017. Through the looking glass: The effects of feedback on self-awareness and conversational behaviour during video chat. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 5271–5283.

[46] William R Miller and Theresa B Moyers. 2006. Eight stages in learning motivational interviewing. *Journal of Teaching in the Addictions* 5, 1 (2006), 3–17.

[47] William R Miller, Theresa B Moyers, Denise Ernst, and Paul Amrhein. 2003. Manual for the motivational interviewing skill code (MISC). *Unpublished manuscript. Albuquerque: Center on Alcoholism, Substance Abuse and Addictions,*

University of New Mexico (2003).

[48] William R Miller and Stephen Rollnick. 2012. *Motivational interviewing: Helping people change.* Guilford press.

[49] William R Miller, R Gayle Sovereign, and Barbara Krege. 1988. Motivational interviewing with problem drinkers: II. The Drinker's Check-up as a preventive intervention. *Behavioural and Cognitive Psychotherapy* 16, 4 (1988), 251–268.

[50] Kátia Monteiro. 2014. An experimental study of corrective feedback during video-conferencing. *Language Learning & Technology* 18, 3 (2014), 56–79.

[51] Hanh Thi Nguyen. 2007. Rapport building in language instruction: A microanalysis of the multiple resources in teacher talk. *Language and Education* 21, 4 (2007), 284–303.

[52] Jennifer Nicolai, Ralf Demmel, and Jutta Hagen. 2007. Rating scales for the assessment of empathic communication in medical interviews (REM): Scale development, reliability, and validity. *Journal of Clinical Psychology in Medical Settings* 14, 4 (2007), 367–375.

[53] Eyal Ofek, Shamsi T. Iqbal, and Karin Strauss. 2013. Reducing Disruption from Subtle Information Delivery During a Conversation: Mode and Bandwidth Investigation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13).* 3111–3120.

[54] Stefan Olafsson, Teresa O'Leary, and Timothy Bickmore. 2019. Coerced change-talk with conversational agents promotes confidence in behavior change. In *Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare.* 31–40.

[55] Stefan Olafsson, Teresa K O'Leary, and Timothy W Bickmore. 2020. Motivating Health Behavior Change with Humorous Virtual Agents. In *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents.* 1–8.

[56] Reza Rawassizadeh, Timothy J. Pierson, Ronald Peterson, and David Kotz. 2018. NoCloud: Exploring Network Disconnection through On-Device Data Analysis. *IEEE Pervasive Computing* 17, 1 (2018), 64–74.

[57] Reza Rawassizadeh, Martin Tomitsch, Manouchehr Nourizadeh, Elaheh Momeni, Aaron Peery, Liudmila Ulanova, and Michael Pazzani. 2015. Energy-Efficient Integration of Continuous Context Sensing and Prediction into Smartwatches. *Sensors* 15, 9 (2015), 22616–22645.

[58] Jeffery R. Ringer. 1985. Pardon Me, Can I Talk Now?: A Look at the Roles of Interruptions in Conversation. *Annual Meeting of the Organization for the Study of Communication, Language, and Gneder* (1985), 1–29.

[59] Stephen Rollnick and William R Miller. 1995. What is motivational interviewing? *Behavioural and cognitive Psychotherapy* 23, 4 (1995), 325–334.

[60] Samiha Samrose. 2018. Automated Collaboration Coach for Video-conferencing based Group Discussions. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers.* 510–515.

[61] Samiha Samrose, Kavya Anbarasu, Ajjen Joshi, and Taniya Mishra. 2020. Mitigating Boredom Using An Empathetic Conversational Agent. In *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents.* 1–8.

[62] Samiha Samrose, Wenyi Chu, Carolina He, Yuebai Gao, Syeda Sarah Shahrin, Zhen Bai, and Mohammed Ehsan Hoque. 2019. Visual cues for disrespectful conversation analysis. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII).* IEEE, 580–586.

[63] Samiha Samrose, Daniel McDuff, Robert Sim, Jina Suh, Kael Rowan, Javier Hernandez, Sean Rintel, Kevin Moynihan, and Mary Czerwinski. 2021. MeetingCoach: An Intelligent Dashboard for Supporting Effective & Inclusive Meetings. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.* 1–13.

[64] Samiha Samrose, Reza Rawassizadeh, and Ehsan Hoque. 2020. Immediate or Reflective?: Effects of Real-timeFeedback on Group Discussions over Videochat. *arXiv preprint arXiv:2011.06529* (2020).

[65] Samiha Samrose, Ru Zhao, Jeffery White, Vivian Li, Luis Nova, Yichen Lu, Mohammad Rafayet Ali, and Mohammed Ehsan Hoque. 2018. Coco: Collaboration coach for understanding team dynamics during video conferencing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (2018), 1–24.

[66] Björn Schuller, Stefan Steidl, Anton Batliner, Elmar Nöth, Alessandro Vinciarelli, Felix Burkhardt, Rob Van Son, Felix Weninger, Florian Eyben, Tobias Bocklet, et al. 2015. A survey on perceived speaker traits: Personality, likability, pathology, and the first challenge. *Computer speech & language* 29, 1 (2015), 100–131.

[67] Daniel Schulman, Timothy W Bickmore, and Candace L Sidner. 2011. An Intelligent Conversational Agent for Promoting Long-Term Health Behavior Change Using Motivational Interviewing.. In *AAAI Spring Symposium: AI and Health Communication.*

[68] Abigail J Sellen. 1995. Remote conversations: The effects of mediating talk with technology. *Human-computer interaction* 10, 4 (1995), 401–444.

[69] David K Sherman and Geoffrey L Cohen. 2002. Accepting threatening information: Self–Affirmation and the reduction of defensive biases. *Current directions in psychological science* 11, 4 (2002), 119–123.

[70] David K Sherman and Geoffrey L Cohen. 2006. The psychology of self-defense: Self-affirmation theory. *Advances in experimental social psychology* 38 (2006), 183–242.

[71] Susan G Straus, Jeffrey A Miles, and Laurie L Levesque. 2001. The effects of videoconference, telephone, and face-to-face media on interviewer and applicant judgments in employment interviews. *Journal of management* 27, 3 (2001), 363–381.

[72] Harry Stack Sullivan. 1973. *Clinical studies in psychiatry.* Vol. 2. WW Norton & Company.

[73] Deborah Tannen. 1997. Gender and Discourse. *Discourse* (1997), 548–567. https://doi.org/10.2307/2655317

[74] M.I. Tanveer, E. Lin, and M.E. Hoque. 2015. Rhema: A Real-time in-situ Intelligent Interface to Help People with Public Speaking. In *Proceedings of the 20th International Conference on Intelligent User Interfaces.* 286–295.

[75] Yla R Tausczik and James W Pennebaker. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology* 29, 1 (2010), 24–54.

[76] Yla R Tausczik and James W Pennebaker. 2013. Improving teamwork using real-time language feedback. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* 459–468.

[77] Janet H Walker, Lee Sproull, and R Subramani. 1994. Using a human face in an interface. In *Proceedings of the SIGCHI conference on human factors in computing systems.* ACM, 85–91.

[78] Joseph B Walther and Lisa C Tidwell. 1995. Nonverbal cues in computer-mediated communication, and the effect of chronemics on relational communication. *Journal of Organizational Computing and Electronic Commerce* 5, 4 (1995), 355–378.

[79] Darrell M West. 2017. *Air wars: Television advertising and social media in election campaigns, 1952-2016.* CQ Press.