Face Value? Exploring the Effects of Embodiment for a Group Facilitation Agent

Ameneh Shamekhi[1*],Q. Vera Liao[2], Dakuo Wang[2], Rachel K. E. Bellamy[2], Thomas Erickson[2]
[1]Northeastern University, Boston, MA, USA
[2]IBM Research, Yorktown Heights, NY, USA
ameneh@ccs.neu.edu, {vera.liao, dakuo.wang}@ibm.com, {rachel, snowfall}@us.ibm.com

ABSTRACT

We are 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. We 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, we 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.

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous; See http://acm.org/about/class/1998/ for the full list of ACM classifiers. This section is required.

Author Keywords

Group Decision Making; Conversational Agents; Meeting Facilitation; Agent's Embodiment

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 interactions 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 [34], agenda setting [36], and preference elicitation [14]) are the targets of current research and technology development. We think the time is ripe to consider the key design issues for group facilitation agents. In this paper, we 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. [9], 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 [57, 25, 26]. Empirical evidence about the necessity of the embodiment is mixed. Some literature [23, 4, 54] 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 [18, 62]. There seems to be a pattern that for tasks that require continuous engagement (e.g., tutoring), embodiment of the agent improves task performance [1, 40]; 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" [25, 56]. 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, be-

* This work was done during an internship at IBM Research

CHI 2018, April 21–26, 2018, Montreal, QC, Canada

DOI: https://doi.org/10.1145/3173574.3173965

cause enforcing structure implies attention and compliance, a more socially favorable agent may be more effective in improving group processes and outcomes. Secondly, affective benefit is often emphasized in technologies supporting collaborative work because it can improve social and collaborative process [37], and positive affect brought by CAs has been observed in casual social settings [50]. Furthermore, we postulate that embodiment may create a stronger sense of presence as in continuously “being there”, especially in a group setting. 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 our study to explore impacts of the agent’s embodiment on group facilitation tasks. We designed and developed an embodied conversational agent that facilitates a group decision task by enforcing a structured discussion process. We 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, we 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. We complemented our quantitative results with qualitative focus group interview data. Our 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 paper, we will first review related work that motivated our study. We then present how we 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 we 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.

RELATED WORK

Our study is mainly informed by two sets of literature: the work on designing computer-supported systems to facilitate group decision-making tasks, and the work on conversational agent design, in particular, the design of agents’ embodiment and its impact on user perceptions and behaviors.

Group Decision Support Systems

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 [49], visualization [28], and group recommender systems. In a 1987 foundational paper, DeSanctics and Gallupe [19] defined 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).

Within the HCI community, the studies of GDSSs can be traced back to the 1980s. Poole et al. [49] 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 groups that used a Level 1 GDSS, with a display of problem definition, criteria for evaluation, alternative solutions and the current group voting status. The work suggested seven types of benefits that GDSSs could provide for group decision-making, 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.

Much research suggests that having a group facilitator can improve group decision-making effectiveness, because a facilitator can enforce a more structured process [53, 58, 38]. Thus, a number of GDSSs were designed to support or even replace some of a human facilitator’s roles [45, 21, 15]. For example, in distributed groups, a lack of leadership is often found to inhibit the groups’ capability to organize and reach consensus [27]. To tackle this problem, Farnham et al. [20] customized a chat system to enforce a more optimal structure of a decision-making process. This system is designed specifically for a hiring decision making scenario to select a best job candidate, where the user groups were instructed to follow three steps: 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 [20].

In this work, we are interested in developing a conversational agent to support group decision making tasks, by providing the agent capabilities to enforce a more structured process. In contrast to [20], we focus on group decision making that happens at the same time and in the same place. According to McGrath’s typology [35], 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 [20], our agent system is designed to imitate a human facilitator by providing pro-active and humanized facilitation. Our agent should be considered a Level 1 GDSS, according to [19]. However, we expect an interactive agent system to be able to support not only the decision making process (production), but also positively influence the interactions of group members (well-being and member support). To explore these effects, we conducted a comparative experiment similar to [20] using a group decision making task of selecting the best job candidate.

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 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. We consider a group facilitation agent to belong to the latter category.

Goal-oriented CAs are used in various domains such as health care [6] and education [55]. For example, Graesser et al. developed AutoTutor, an animated pedagogical agent [22] and demonstrated that the agent significantly improved learning outcomes by engaging students in an interactive conversation [55]. 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 pedagogical agents to enable them to handle complex interaction dynamics in a multi-learner collaborative environment [31]. Bohus and Horvitz also proposed a model to represent a CA's turn-taking behaviors by gaze, gestures and speech in multiparty conversations [7].

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 [11, 10]. Therefore, ECA can serve as an interface for more intuitive and engaging interactions [12]. 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. [32]. 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 [44], motivation [3], entertainment [30] and trust [4]. 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[59]. 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 [39, 1, 40].

On the negative side, Hasegawa et al. studied a direction-giving service provided by an ECA and a GPS system, [25] 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 [26]. 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, we 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. We aim to contribute both to practical guidelines for designing CAs in collaborative contexts, and new knowledge on the effects of embodiment.

RESEARCH METHOD

We developed a conversational agent and designed a Wizard-of-Oz experiment[17] 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, we first describe the experiment task and design of the CA. We then present the facilitation protocol of the agent. Lastly we lay out the details of the experiment design.

Experiment Task

We 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., [20]). 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. We 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.

Embodied Conversational Agent Design

We designed an embodied conversational agent (ECA) named CASSY (Collaborative Agent for Decision Support System) to facilitate group decision-making sessions. We 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[1], both to fit a face-area representation and create a sense of mobility to enhance her presence (Figure 1).

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 [13]).

In this study, we 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 [4, 60, 8]. 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. We used a constrained Wizard-of-Oz protocol [51] 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).

[1]Manufactured by Suitable Technologies Beam Pro: http://suitabletech.com

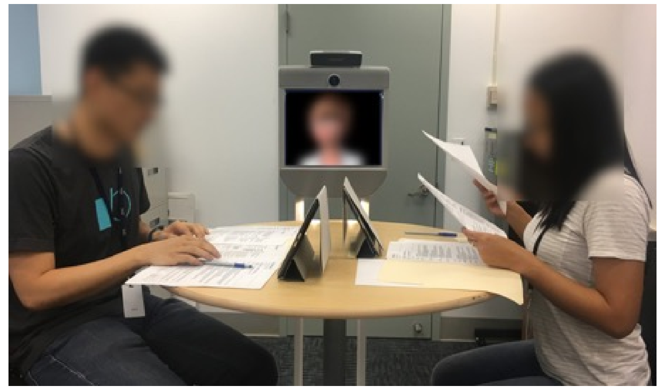

Figure 1. Experiment setting: CASSY projected on a Beam telepresence robot, and facilitating a group decision making task with two participants. In front of the participants, there are job description, rating sheet, and resumes. Two iPads on the table serve as video recording devices.

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 understanding of the problem, 2) full exploration of each alternative, and 3) comparison of alternatives [63]. We 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 provides 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 understanding 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).

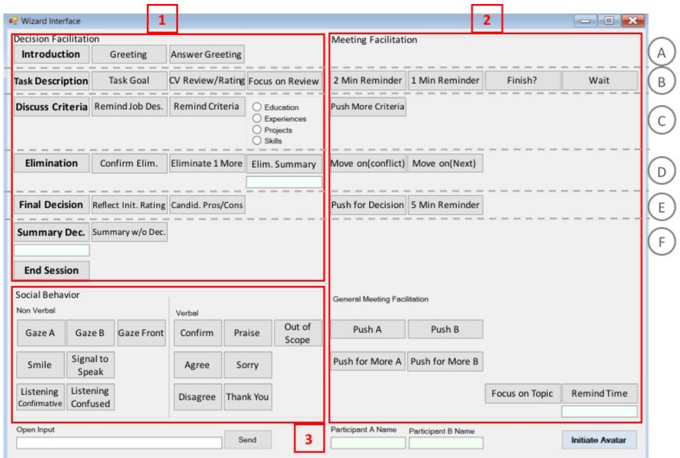

Figure 2. The wizard interface. 1-3 shows the three forms of agent's services, and A-F shows the six states of each decision making session.

Below we present a procedural view of the protocol, which covers most of the decision facilitation and meeting facilitation the wizard provides. We describe how the wizard selects intents on the Wizard interface (Figure 2). 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.

1. *Introduction*: CASSY initiates the conversation by introducing herself, inviting the participants to introduce themselves, and greeting them with their names. (Figure 2-A).
2. *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 (Figure 2-B).
3. *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 discussed. 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. (Figure2-C)
4. *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 can not 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. (Figure2-D)
5. *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 (Figure2-E).
6. *Exit the Experiment*: CASSY summarizes the session, either with a final decision or without a consensus, and thanks the group for participating. (Figure2-F).

Experimental design and procedure

We 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 (Figure 1); in the voice condition (no Beam), CASSY's voice was played via the same speaker on the table. Figure 1 shows 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, we conducted 20-30 minute semi-structured focus groups with our participants. We started by asking questions about their overall impression about the decision process and the agent's facilitation. We 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 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 that 35. To minimize the effect of gender and expertise on the group performance, we 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 $p = 1$). We 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

Our measurements consist of subjective responses from a survey, and objective measures reflecting the decision process and outcome. In this section we 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, we asked participants to rate the agent's rapport, power, anthropomorphism, intelligence, and trust. To complement the objective measures on group processes, we asked participants to report subjective evaluation of their experience with the decision making process, and towards their collaborator. We also collected demographic information such as gender and race at the end of the survey.

User Perceptions of the Agent

To answer RQ1, we 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], we 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"*.

We 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. We adapted validated scales on these dimension presented by [2] and [46] by selecting 2 items for each scale. We 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, we 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). We 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, we 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. We also asked participants' subjective evaluation of the group decision performance in the survey. As suggested by [24] we 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] we 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 our quantitative analysis is to compare participants' ratings of agent perceptions between the two conditions—avatar v.s. 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, we 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). We 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 we used. We also tested our 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 our exploratory analysis showed they yielded the same conclusions as LMMs. In the rest of the paper, we report results from LMM, whenever the random group effect needs to be considered. We ran LMMs with the *lme4* package of R, with the default restricted maximum likelihood estimation. To report p-values of variable effects, we used likelihood ratio tests (R procedures as seen in [61]), a common approach for significance testing of LMMs

RESULTS

In the following sections, we 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, we present our qualitative findings and further discuss the pros and cons of agent's embodiment in the group context (RQ1 and RQ2).

Agent perceptions (RQ1)

To answer RQ1, we examined participants' subjective perceptions of CASSY based on their survey responses. Specifically, we 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 methodology. For each scale, we conducted a linear mixed model regression by including the condition (voice vs. avatar) as a fixed factor, and group ID as a random factor. We removed one group that appeared to be an outlier with multiple agent perception scales. In Table 1, we 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 [2]. For RQ1, we 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.

Decision performance (RQ2-a)

Next we 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 we 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). We examined two kinds of rating shift: 1) Consensus shift, for which we 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

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

Agent Scale	Descriptive statistics M_{Avatar} (SD)	M_{voice} (SD)	LMM statistics for fixed effect (condition) (intercept)	β	SE	CI (95%)	logLik	Chi -square	p -value	Random effect (group) Variance Est.	Std. 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†

Table 1. 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.

between -1 (no agreement at all) to 1 (complete agreement). 2) Individual shift, for which we calculated the ICC of each participant's pre and post ratings. So a higher individual ICC indicated less individual opinion shift. We 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.

We also looked at the groups' top choices. All groups but one reached agreement on the top candidate. While we did not intentionally design the study with a best choice, we 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, we 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 2.

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

Group interactions (RQ2-b)

We further explored whether the embodiment had any effect on interactions between the group members. By looking into the survey items on perception of partners, we 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" [29]. 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.

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	χ^2	1
Confidence improve	1.1 (0.79)	1.3 (0.98)	LMM[5]	0.49
Time	21.6 (4.72)	22.8 (3.55)	TT	0.53

Table 2. Results of decision outcome measures (TT:T-test, MM:mixed model linear regression. $^4\beta = 0.04, SE = 0.143$, $^5\beta = -0.2, SE = 0.289$

To test this hypothesis, we compared the contribution equality of pairs between the two conditions. After transcribing all discussions, we counted the number of turns and the number of words from each participant through a text analytic process. We 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, we 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.

Agent interactions (RQ2-c)

Lastly, we looked into participants' interactions with the agent. Our 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 we 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, we 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 3. In total, proactive interactions more than doubled in the groups with embodied agents and one tailed t-test showed marginal significance

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)

Table 3. Average occurrences of proactive interactions with the agent; standard deviations are in the parentheses

($t(18) = 1.6015$, $p = 0.06$). We conclude that the embodiment invited more proactive interactions from the participants.

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, we 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, we reflect on the implications of embodiment for the application domain. All focus-group interviews were transcribed and coded using thematic analysis techniques. The results of our 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. We 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" [29]. We 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 unable 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 = 5.75, SD = 1.61) compared to the voice condition (M = 4.95, SD = 1.93, $p = 0.16$, not statistically significant but indicating a trend).

In the survey responses, we 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, we 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 confirmed 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 we cannot reach an agreement and then we 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. We speculate that the keys lie in attention and compliance, especially when the agent suggests actions that are contrary to the individuals' current engagement. In our 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. Meanwhile, 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.

We 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 we found embodiment could potentially increase a sense of power, we 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. We 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 [19]. 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 [10].

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, we discuss how our results—particularly our 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

Our study showed positive influence of embodiment on how users perceive and respond to a conversational agent in a group context. Our 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". We discuss some possible designs to further enhance these two aspects, specifically for group facilitation agents.

In our design of embodied agents, and in many previous studies (e.g., [5]), 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, we 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 follow a protocol to handle the interruptions (e.g. immediate stop [12], 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 consider 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) exhibit 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 informational 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 [41], and creating a mismatch between user expectations and system capabilities [33]. Our 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 our study on the benefit of embodiment is enhanced presence. Its implications merit particular consideration for introducing CAs into collaborative roles. As some of our 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.

Our 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.

Limitations
We acknowledge several limitations of the study: 1) we could only conjecture about some of the causalities for the agent's effect on group processes. The validation awaits future research. 2) The "hiring decision" experimental task may impact effects we 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. We 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, we controlled for factors that could impact how groups react to a facilitation agent, for example, the number of participants and power dynamics. Our 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. We acknowledge that the setup may have not observed certain group phenomena. We hope future research explore how embodiment interacts with different individual, social and cultural factors.

Conclusion
We revisited the longstanding debate on the necessity of embodiment for conversational agents in a new context–group facilitation. Consistent with previous work, we 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, we found evidence that the embodiment had positive influence on group dynamics and invited more pro-active interactions. Our 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, we 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.

ACKNOWLEDGMENT
We would like to thank all study participants, and Victor Dibia, Jason Ellis, Jeffrey Kephart, Nicholas Mattei, Kartik Talamadupula, Biplav Srivastava, and Yunfeng Zhang for their valuable feedback for the study design.

REFERENCES

1. Robert K Atkinson. 2002. Optimizing learning from examples using animated pedagogical agents. *Journal of Educational Psychology* 94, 2 (2002), 416.

2. Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics* 1, 1 (2009), 71–81.

3. Amy L Baylor. 2009. Promoting motivation with virtual agents and avatars: role of visual presence and appearance. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 364, 1535 (2009), 3559–3565.

4. Timothy Bickmore and Justine Cassell. 2001. Relational agents: a model and implementation of building user trust. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 396–403.

5. Timothy Bickmore and Justine Cassell. 2005. Social dialogue with embodied conversational agents. *Advances in natural multimodal dialogue systems* 30 (2005), 23–54.

6. Timothy W Bickmore, Laura M Pfeifer, and Brian W Jack. 2009. Taking the time to care: empowering low health literacy hospital patients with virtual nurse agents. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 1265–1274.

7. Dan Bohus and Eric Horvitz. 2010. Facilitating multiparty dialog with gaze, gesture, and speech. In *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction*. ACM, 5.

8. Jay Bradley, David Benyon, Oli Mival, and Nick Webb. 2010. Wizard of Oz experiments and companion dialogues. In *Proceedings of the 24th BCS Interaction Specialist Group Conference*. British Computer Society, 117–123.

9. Justine Cassell. 2000. *Embodied conversational agents*. MIT press.

10. Justine Cassell. 2001. Embodied conversational agents: representation and intelligence in user interfaces. *AI magazine* 22, 4 (2001), 67.

11. Justine Cassell and Timothy Bickmore. 2000. External manifestations of trustworthiness in the interface. *Commun. ACM* 43, 12 (2000), 50–56.

12. Justine Cassell, Timothy Bickmore, Mark Billinghurst, Lee Campbell, Kenny Chang, Hannes Vilhjálmsson, and Hao Yan. 1999. Embodiment in conversational interfaces: Rea. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, 520–527.

13. Justine Cassell, Hannes Högni Vilhjálmsson, and Timothy Bickmore. 2001. Beat: the behavior expression animation toolkit. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*. ACM, 477–486.

14. Konstantina Christakopoulou, Filip Radlinski, and Katja Hofmann. 2016. Towards Conversational Recommender Systems.. In *KDD*. 815–824.

15. Peter Cook, Clarence Ellis, Mike Graf, Gail Rein, and Tom Smith. 1987. Project Nick: meetings augmentation and analysis. *ACM Transactions on Information Systems (TOIS)* 5, 2 (1987), 132–146.

16. Duncan Cramer and Dennis Laurence Howitt. 2004. *The Sage dictionary of statistics: a practical resource for students in the social sciences*. Sage.

17. Nils Dahlbäck, Arne Jönsson, and Lars Ahrenberg. 1993. Wizard of Oz studiesâĂŤwhy and how. *Knowledge-based systems* 6, 4 (1993), 258–266.

18. Doris M Dehn and Susanne Van Mulken. 2000. The impact of animated interface agents: a review of empirical research. *International journal of human-computer studies* 52, 1 (2000), 1–22.

19. Gerardine Desanctis and R Brent Gallupe. 1987. A Foundation For The Study Of Group Decision Support Systems. *Management Science* 33 (1987), N0.

20. Shelly Farnham, Harry R Chesley, Debbie E McGhee, Reena Kawal, and Jennifer Landau. 2000. Structured online interactions: improving the decision-making of small discussion groups. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work*. ACM, 299–308.

21. Gregg Foster and Mark Stefik. 1986. Cognoter: theory and practice of a colab-orative tool. In *Proceedings of the 1986 ACM conference on Computer-supported cooperative work*. ACM, 7–15.

22. Arthur C Graesser, Patrick Chipman, Brian C Haynes, and Andrew Olney. 2005. AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions on Education* 48, 4 (2005), 612–618.

23. Jonathan Gratch, Ning Wang, Jillian Gerten, Edward Fast, and Robin Duffy. 2007. Creating rapport with virtual agents. In *International Workshop on Intelligent Virtual Agents*. Springer, 125–138.

24. Michael Haller, Jakob Leitner, Thomas Seifried, James R Wallace, Stacey D Scott, Christoph Richter, Peter Brandl, Adam Gokcezade, and Seth Hunter. 2010. The nice discussion room: Integrating paper and digital media to support co-located group meetings. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 609–618.

25. Dai Hasegawa, Justine Cassell, and Kenji Araki. 2010. The Role of Embodiment and Perspective in Direction-Giving Systems.. In *Aaai fall symposium: Dialog with robots*.

26. Renate Häuslschmid, Max von Buelow, Bastian Pfleging, and Andreas Butz. 2017. SupportingTrust in Autonomous Driving. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces*. ACM, 319–329.

27. Starr Roxanne Hiltz, Kenneth Johnson, and Murray Turoff. 1986. Experiments in group decision making communication process and outcome in face-to-face versus computerized conferences. *Human communication research* 13, 2 (1986), 225–252.

28. Joshua E Introne. 2009. Supporting group decisions by mediating deliberation to improve information pooling. In *Proceedings of the ACM 2009 international conference on Supporting group work*. ACM, 189–198.

29. Sam Kaner. 2014. *Facilitator's guide to participatory decision-making*. John Wiley & Sons.

30. Tomoko Koda and Pattie Maes. 1996. Agents with faces: The effect of personification. In *Robot and Human Communication, 1996., 5th IEEE International Workshop on*. IEEE, 189–194.

31. Rohit Kumar and Carolyn P Rose. 2011. Architecture for building conversational agents that support collaborative learning. *IEEE Transactions on Learning Technologies* 4, 1 (2011), 21–34.

32. James C Lester, Sharolyn A Converse, Susan E Kahler, S Todd Barlow, Brian A Stone, and Ravinder S Bhogal. 1997. The persona effect: affective impact of animated pedagogical agents. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*. ACM, 359–366.

33. Ewa Luger and Abigail Sellen. 2016. Like having a really bad PA: the gulf between user expectation and experience of conversational agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 5286–5297.

34. Yoichi Matsuyama, Iwao Akiba, Shinya Fujie, and Tetsunori Kobayashi. 2015. Four-participant group conversation: A facilitation robot controlling engagement density as the fourth participant. *Computer Speech & Language* 33, 1 (2015), 1–24.

35. Joseph E McGrath. 1991. Time, interaction, and performance (TIP) A Theory of Groups. *Small group research* 22, 2 (1991), 147–174.

36. Moira McGregor and John C Tang. 2017. More to Meetings: Challenges in Using Speech-Based Technology to Support Meetings.. In *CSCW*. 2208–2220.

37. Nicolas Michinov and Estelle Michinov. 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 (2008), 1540–1557.

38. Shalla M Miranda and Robert P Bostrom. 1999. Meeting facilitation: process versus content interventions. *Journal of Management information systems* 15, 4 (1999), 89–114.

39. Roxana Moreno, Richard Mayer, and James Lester. 2000. Life-like pedagogical agents in constructivist multimedia environments: Cognitive consequences of their interaction. In *EdMedia: World Conference on Educational Media and Technology*. Association for the Advancement of Computing in Education (AACE), 776–781.

40. Roxana Moreno, Richard E Mayer, Hiller A Spires, and James C Lester. 2001. The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and instruction* 19, 2 (2001), 177–213.

41. Masahiro Mori. 1970. The uncanny valley. *Energy* 7, 4 (1970), 33–35.

42. Geoff Norman. 2010. Likert scales, levels of measurement and the âĂIJlawsâĂİ of statistics. *Advances in health sciences education* 15, 5 (2010), 625–632.

43. David Novick and Iván Gris. 2014. Building rapport between human and ECA: A pilot study. In *International Conference on Human-Computer Interaction*. Springer, 472–480.

44. Kristine L Nowak and Frank Biocca. 2003. The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators and Virtual Environments* 12, 5 (2003), 481–494.

45. Jay F. Nunamaker, Alan R Dennis, Joseph S Valacich, Douglas Vogel, and Joey F George. 1991. Electronic meeting systems. *Commun. ACM* 34, 7 (1991), 40–61.

46. Jay F Nunamaker, Douglas C Derrick, Aaron C Elkins, Judee K Burgoon, and Mark W Patton. 2011. Embodied conversational agent-based kiosk for automated interviewing. *Journal of Management Information Systems* 28, 1 (2011), 17–48.

47. Amy Ogan, Samantha L Finkelstein, Erin Walker, Ryan Carlson, and Justine Cassell. 2012. Rudeness and Rapport: Insults and Learning Gains in Peer Tutoring.. In *ITS*. Springer, 11–21.

48. José C Pinheiro and Douglas M Bates. 2000. Linear mixed-effects models: basic concepts and examples. *Mixed-effects models in S and S-Plus* (2000), 3–56.

49. Marshall Scott Poole, Michael Homes, and Gerardine DeSanctis. 1988. Conflict management and group decision support systems. In *Proceedings of the 1988 ACM conference on Computer-supported cooperative work*. ACM, 227–243.

50. Martin Porcheron, Joel E Fischer, and Sarah Sharples. 2016. Using mobile phones in pub talk. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 1649–1661.

51. Laurel D Riek. 2012. Wizard of oz studies in hri: a systematic review and new reporting guidelines. *Journal of Human-Robot Interaction* 1, 1 (2012).

52. Eduardo Salas, Rebecca Grossman, Ashley M Hughes, and Chris W Coultas. 2015. Measuring team cohesion: Observations from the science. *Human factors* 57, 3 (2015), 365–374.

53. Gregory P Sbea and Richard A Guzzo. 1987. Group effectiveness: what really matters? *Sloan Management Review (1986-1998)* 28, 3 (1987), 25.

54. Akikazu Takeuchi and Taketo Naito. 1995. Situated facial displays: towards social interaction. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM Press/Addison-Wesley Publishing Co., 450–455.

55. Ronald K Thornton and David R Sokoloff. 1998. Assessing student learning of NewtonâĂŹs laws: The force and motion conceptual evaluation and the evaluation of active learning laboratory and lecture curricula. *American Journal of Physics* 66, 4 (1998), 338–352.

56. Susanne Van Mulken, Elisabeth André, and Jochen Müller. 1998. The persona effect: How substantial is it? In *People and computers XIII*. Springer, 53–66.

57. Susanne Van Mulken, Elisabeth André, and Jochen Müller. 1999. An empirical study on the trustworthiness of life-like interface agents.. In *HCI (2)*. 152–156.

58. Stephen Viller. 1991. The group facilitator: a CSCW perspective. In *Proceedings of the Second European Conference on Computer-Supported Cooperative Work ECSCWâĂŹ91*. Springer, 81–95.

59. 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.

60. Pierre Wargnier, Giovanni Carletti, Yann Laurent-Corniquet, Samuel Benveniste, Pierre Jouvelot, and Anne-Sophie Rigaud. 2016. Field evaluation with cognitively-impaired older adults of attention management in the embodied conversational agent louise. In *Serious Games and Applications for Health (SeGAH), 2016 IEEE International Conference on*. IEEE, 1–8.

61. Bodo Winter. 2013. A very basic tutorial for performing linear mixed effects analyses. *arXiv preprint arXiv:1308.5499* (2013).

62. Nick Yee, Jeremy N Bailenson, and Kathryn Rickertsen. 2007. A meta-analysis of the impact of the inclusion and realism of human-like faces on user experiences in interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 1–10.

63. Alvin Frederick Zander. 1982. *Making groups effective*. Jossey-Bass.

64. Ran Zhao, Alexandros Papangelis, and Justine Cassell. 2014. Towards a dyadic computational model of rapport management for human-virtual agent interaction. In *International Conference on Intelligent Virtual Agents*. Springer, 514–527.