Check for updates

How Should the Agent Communicate to the Group? Communication Strategies of a Conversational Agent in Group Chat Discussions

HYO JIN DO, University of Illinois at Urbana-Champaign, USA
HA-KYUNG KONG, Seattle University, USA
JAEWOOK LEE, University of Illinois at Urbana-Champaign, USA
BRIAN P. BAILEY, University of Illinois at Urbana-Champaign, USA

387

In online group discussions, balanced participation can improve the quality of discussion, members' satisfaction, and positive group dynamics. One approach to achieve balanced participation is to deploy a conversational agent (CA) that encourages participation of under-contributing members, and it is important to design communication strategies of the CA in a way that is supportive to the group. We implemented five communication strategies that a CA can use during a decision-making task in a small group synchronous chat discussion. The five strategies include messages sent to two types of recipients (@username vs. @everyone) crossed by two separate channels (public vs. private), and a peer-mediated strategy where the CA asks a peer to address the under-contributing member. Through an online study with 42 groups, we measured the balance of participation and perceptions about the CA by analyzing chat logs and survey responses. We found that the CA sending messages specifying an individual through a private channel is the most effective and preferred way to increase participation of under-contributing members. Participants also expressed that the peer-mediated strategy is a less intrusive and less embarrassing way of receiving the CA's messages compared to the conventional approach where the CA directly sends a message to the under-contributing member. Based on our findings, we discuss trade-offs of various communication strategies and explain design considerations for building an effective CA that adapts to different group dynamics and situations.

CCS Concepts: • **Human-centered computing → Empirical studies in HCI**; **Collaborative interaction**; *Natural language interfaces.*

Additional Key Words and Phrases: Conversational Agent, Collaborative Task, Participation, Group Chat

ACM Reference Format:
Hyo Jin Do, Ha-Kyung Kong, Jaewook Lee, and Brian P. Bailey. 2022. How Should the Agent Communicate to the Group? Communication Strategies of a Conversational Agent in Group Chat Discussions. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 387 (November 2022), 23 pages. https://doi.org/10.1145/3555112

1 INTRODUCTION

Balanced participation is one of the crucial aspects of online group discussions that leads to higher satisfaction [57] and performance [20]. However, low participating behavior of one or few members is common in discussions [36] and online chats are no exceptions [49]. To address the issue,

Authors' addresses: Hyo Jin Do, hjdo2@illinois.edu, University of Illinois at Urbana-Champaign, Urbana, IL, USA; Ha-Kyung Kong, hkong@seattleu.edu, Seattle University, P.O. Box 1212, Seattle, WA, USA; Jaewook Lee, jaewook4@cs.washington.edu, University of Illinois at Urbana-Champaign, Urbana, IL, USA; Brian P. Bailey, bpbailey@illinois.edu, University of Illinois at Urbana-Champaign, Urbana, IL, USA.

2573-0142/2022/11-ART387 $15.00
https://doi.org/10.1145/3555112

conversational agents (CAs) are increasingly deployed and tested to engage under-contributing members in online group discussions [18, 58, 64]. CAs allow greater scalability than human moderators and can make timely interventions using natural language processing techniques. In this work, we design a CA that effectively communicates with under-contributing members by giving them an opportunity to contribute, and increasing social pressure and accountability to participate.

The effectiveness of CAs can vary depending on how messages are communicated. Tegos et al. tested two types of communication strategies, an individual strategy directly specifying an individual (e.g., "Janna, can you give an example...") and a group strategy addressing the whole group (e.g., "Can you give an example..."). The authors found that the individual strategy can increase a productive dialogue between members, but is less preferred by users than the group strategy [58]. The authors tested these strategies in a public channel where CA messages are visible to all members, but how these messages compare in a private channel where only the recipient can read these messages is unknown. Users may prefer the individual over the group strategy when using a private channel because they do not feel normative pressure or surveillance from other members [52, 67]. On the other hand, these ways of communication assume that a CA makes interventions directly to an under-contributing member. Walker proposed that peer-mediated support in which peers process and deliver systems' feedback to the tutee can promote deeper interaction between the students [61]. Yet no study has applied this concept to a CA to encourage an under-contributing member through peers. This work extends the research on communication strategies of a CA to understand the effects of individual and group strategies in private and public channels and the peer-mediated strategy on balanced participation.

We conducted a 2 x 2 factorial experiment where we manipulated the *specificity* of a message by comparing individual versus group recipients – a message can specify an individual (e.g., "@username") or address the whole group (e.g., "@everyone") – crossed by the *publicness* of a message by using public and private channels – everyone can see the public messages whereas only the receiver can see the private messages. Additionally, we tested a *peer-mediated* strategy, where the CA prompts another member of the group to address an under-contributing member publicly (e.g., "@peer, would you please involve @username in your conversation?"). We tested these strategies through an online study where participants carried out a collaborative decision-making task through a synchronous chat discussion. A CA was deployed in chat rooms that messaged the least participating member for opinions using one of the communication strategies. After the discussion, participants completed a questionnaire individually. Through chat logs and survey responses, we measured members' participation, perceptions of the CA, and decision outcomes.

The contributions of this study are threefold. First, we identified that the private individual strategy is the most effective communication strategy overall to balance participation as well as to reduce the intrusiveness of the CA and embarrassment, compared to public and group strategies. We also found that the peer-mediated strategy can improve perceptions of the agent compared to the conventional strategy where a CA directly prompts an under-contributing member. Second, we reported trade-offs of different communication strategies based on qualitative findings and offered guidance on how each strategy can be used depending on the desired outcome in groups. For example, we recommend public strategies when all members' contributions are low, group strategies if one member is likely to feel singled out by repeated prompts, and the peer-mediated strategy in a classroom discussion where team inclusiveness is crucial. Third, we offer design implications considering various group dynamics such as how to reduce burden of peers when performing the mediator's role. In sum, we demonstrated the importance of communication strategies of a facilitator CA because it can impact members' behaviors and user experiences when working in groups.

2 RELATED WORK

In this section, we first introduce conversational agents and their communication strategies, and then summarize technologies that support balanced participation.

2.1 Communication Strategies of Conversational Agents

Our study designed a conversational agent (CA) for promoting participation of under-contributing members in small group chat discussions. Conversational agents are defined as dialogue systems that converse automatically using natural language. Communication strategies of CAs are under-explored, especially in group settings, compared to abundant research on designing message content [32, 64, 68]. Our study explored three types of communication strategies: specificity, publicness, and peer-mediation. The specificity of a CA message refers to a message specifying an *individual* recipient or addressing the whole *group*. The most common strategy used by CAs is to directly address an individual by tagging one's username (e.g., @username, what do you think?) [23, 32, 64]. Alternatively, the CA can send a message without specifying one's username but rather addressing the whole group (e.g., @everyone, what do you think?). Tegos et al. built a collaborative learning system where a CA sends public prompts for productive learning [58] and compared individual and group strategies. They provided preliminary findings that individual messages specifying a student who needs support can encourage interactions between members compared to messages addressing the whole group. However, students preferred the group strategy more, which authors presumed to be due to the freedom to respond – participants can more freely decide when to respond or whether to ignore the message. The social impact theory supports the use of the individual strategy because the social influence is expected to increase when the number of recipients who receives the impact from the CA decreases [43].

For the publicness, we studied the effects of *private* and *public* channels of communication when delivering facilitation messages from a CA. When sending a message directed to an under-contributing individual, most existing CAs have used public messages that are visible to all members [32, 58, 70] rather than private messages that only the recipient can see. Prior literature on persuasion and social influence endorses the idea of using public channels because messages that are expressed in public can gain higher social influence to change behaviors or attitudes of a person due to normative pressures, compared to private messages where the group is unaware of an individual's responses [67]. However, Schiavo et al. suggested using private channels because a participant may feel uncomfortable as messages publicly reveal their under-participation [52]. Our research was motivated by these studies and evaluated how the publicness and the specificity of a CA message influence members' participation and perception.

The *peer-mediated* strategy that prompts another member in the group to deliver facilitation messages to the least participating member was motivated by the indirect assistance method for collaborative learning in which a peer tutor processes the system's feedback presented on a pop-up window and tailors it to the tutee [62]. The authors found that it activates peer tutors' feelings of efficacy and makes peer tutors perceive the system support is relevant. Seo found that the peer-moderated online discussions lead to more in-depth conversations and active engagement from students than discussions without moderation [53]. In this paper, we attempt to offer preliminary evidence about the opportunity of CAs that communicate with the group indirectly using peer-mediation, compared to conventional strategies where CAs directly intervene in group conversations.

2.2 Technologies to Support Balanced Participation

Various technologies have been developed to achieve balanced participation in online group conversations. Social visualization is one of the well-studied methods used by these systems that visually delineates real-time group dynamics using social proxies [21] so that users can reflect and adjust their behaviors [6, 33, 38, 52, 60]. For example, Leshed et al. implemented Groupmeter that visualized linguistic indicators as bar graphs and animations. The authors found that the visualization helped groups reflect on their language use and change their group behaviors [38]. However, visualizations can cognitively overload or distract the users especially during fast-paced online chats and relies on individuals to figure out on their own how to change their behavior [27].

Another method that has been explored is technology to support human facilitators who encourage the participation of group members in online discussions [10, 37]. For example, Lee et al. implemented the SolutionChat system to assist human facilitators by visualizing discussion stages and recommending contextually appropriate facilitation messages [37]. However, approaches that involve human facilitators have limited scalability, require adequate training, and are costly. To address these limitations, many researchers have also proposed automatic feedback systems to facilitate teamwork [24, 57]. For example, Tausczik and Pennebaker proposed a real-time language feedback system that displays feedback using pop-up windows. However, students perceived the pop-ups distracting and cognitively overloading when presented with multiple feedback messages [57].

Prior studies have found promising results for the usage of CAs that play the role of a facilitator in group conversations [32, 64, 70]. For example, Kim et al. proposed GroupfeedBot which promoted participation of under-contributing members and observed more diversity in opinions [32]. Wang et al. proposed a chat-based agent named Bazaar that supported transactive exchange of ideas and improved quality of the outcome and participants' multi-perspective knowledge [64]. Depending on the context of usage [9], CAs can act as a member in the group rather than a tool [32], offering a more natural, engaging, and less distracting user experience, compared to existing feedback tools (e.g., delivering messages using pop-up windows [57] or displays [52]). This perspective extends the Computers Are Social Actors (CASA) paradigm [46] that people would respond in the same manner regardless of whether they are interacting with a human facilitator or a computer. CAs can create a sense of social presence by adopting human-human communication characteristics [47, 51]. For example, a CA with a humanoid face improved group's social perception of the agent such as rapport, trust, intelligence, and power [54].

We aimed to build a CA that supports balanced participation and minimizes intrusion. Our investigation on communication strategies can be helpful to other CAs deployed in group-based communications such as teaching assistant agents in collaborative learning platforms [63] or online community bots promoting civil group discussions [12], and thereby advance knowledge about agent-team interactions [2]. More broadly, our work contributes to bodies of research on group support systems that deal with various challenges in collaborative works [31] such as resolving group conflicts [28], summarizing and tagging information [70], and reaching consensus [71].

3 METHOD

Our goal was to evaluate different communication strategies for a conversational agent (CA). More specifically, we planned to answer the following research questions:

- **RQ1**. How do the specificity and the publicness of communication strategies interrelate to affect balanced *participation*?
- **RQ2**. How do the specificity and the publicness of communication strategies interrelate to affect user *perception* of the agent in terms of perceived effectiveness, intrusiveness, embarrassment, and social influence?

Table 1. Six experimental conditions. We conducted a between-subject design experiment where one of the six communication strategies was used in each condition.

Communication Strategy	Example Message Prompt
1. Public individual	Hey @username, you have written relatively fewer messages. Your opinion is important because opinions from diverse people can lead to a more creative solution. Would you tell your team your opinions?
2. Private individual	
3. Public group	Hey @everyone, some people wrote relatively fewer messages. Every opinion is important because opinions from diverse people can lead to a more creative solution. Would you tell your team your opinions?
4. Private group	
5. Peer-mediated	Hey @peer, messages from @username were relatively few. Since you seem to like talking, would you please involve @username in your conversation?"
6. Control	Researchers say that balanced participation during group discussions is important. Opinions from diverse people can lead to a more creative solution.

- **RQ3**. How do the specificity and the publicness of communication strategies interrelate to affect group *performance*?
- **RQ4**. How does the *peer-mediation* strategy compare to the public individual strategy in terms of participation, perceptions of the agent, and group performance?

3.1 Study Design

To answer RQ1 to RQ3, we designed a 2 x 2 between-subjects factorial design experiment with two factors, publicness (private, public) and specificity (individual, group). The *publicness* factor indicates the publicness of the agent's communication; messages sent using a public channel are visible to all members, whereas messages sent via a private channel are only visible to the recipient. The *specificity* factor refers to the target specificity of the agent's communication; the individual strategy pinpoints an under-contributing member by mentioning one's username in the message, whereas the group strategy involves everyone in the message. Based on social influence theories [43, 67], we hypothesized that the effects of the specificy on balanced participation vary depending on the publicness of the CA message. For the public channel, the individual strategy could increase the participation of under-contributing members more than the group strategy due to normative pressure and higher social influence from the CA. For the private channel, the individual strategy will be less effective than the group strategy because recipients believe that the other members are unaware that they received messages from the agent and their reactions.

We added a peer-mediated strategy condition to answer RQ4 in which the CA privately asked a peer to involve an under-contributing member in the conversation. We separated RQ4 from the other research questions because our goal was to explore the potential of peer-mediation before investigating interactions with other communication strategies. We designed the peer-mediated strategy as an alternative approach of the conventional public individual strategy; the peer was instructed to address an under-contributing member publicly and specifying the member's username. We chose the public individual strategy for comparison because it is the most common strategy used in prior works (e.g., [18, 32, 62]). Also, it reduces possible confounding factors by allowing members to communicate with others only through the public channel same as the peer-mediated strategy. We also had a control condition in which a CA made a general statement about the importance of balanced participation. We designed this control condition to separate the effects of the proposed communication strategies from the salience bias, which a message itself can predispose individuals to become aware of their participation levels rather than the communication strategy of the message. All experimental conditions are summarized with examples in Table 1.

We conducted an experiment with groups of 3-6 members where everyone is generally expected to participate. We focused on non-hierarchical groups who meet for the first time over text-based synchronous chat. The main reasons for choosing this group setting are: 1) small groups provide a microcosm of any group dynamics and many larger conversations often splinter into smaller

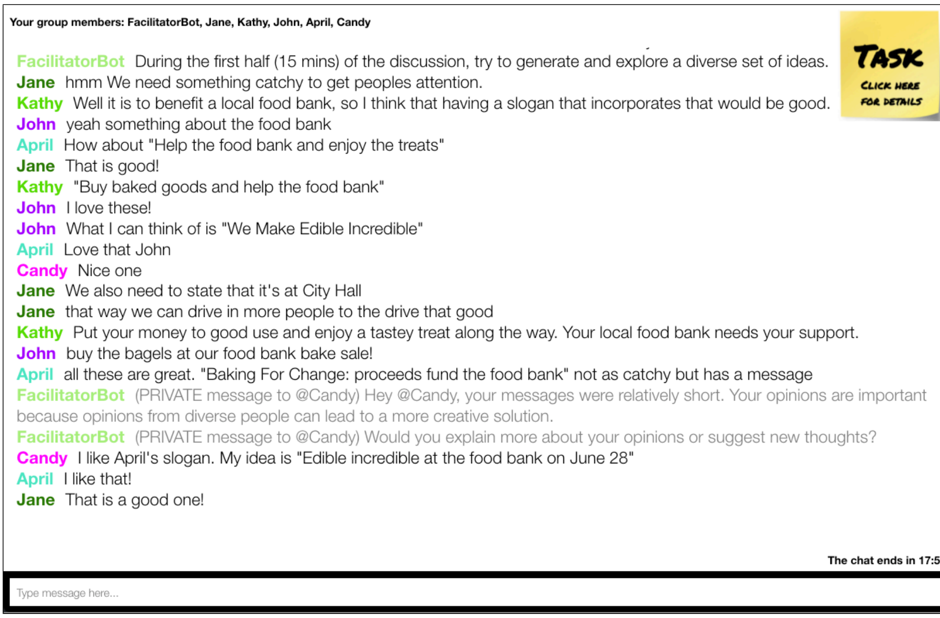

Fig. 1. A text-based chat interface. The interface shows members' usernames, chat history, an input text box, and a timer. Participants can click to the yellow post-it button to read task information again. A CA named FacilitatorBot sends messages to the least contributing members using one of the communication strategies depending on the condition. In this example, FacilitatorBot used the private individual strategy because the message specifies an individual (@Candy) and sent privately to the person only.

groups [6], 2) we wanted to minimize other variables from confounding the effect of the CA such as existing relationship between members, 3) strangers feel more comfortable to talk over chat for the first time rather than video-based or in-person meetings [8], and 4) a synchronous chat is more effective for short discussion compared to asynchronous discussions because it provides a spontaneous dynamic of live conversation [14].

3.2 Task Description

We used an advertising slogan task [1, 17, 41] as an example of decision-making tasks for open-ended problems where a group of people work together online to discuss ideas and develop creative solutions. The task presented a fictional local event, a bake sale fundraiser, and asked participants to discuss and create a one-sentence advertising slogan without a character limit. The details of the event were communicated through a description that was short and easily understandable, but at the same time, gave enough details (e.g., date, location, sale items) to stimulate the creation of diverse slogans. The advertising slogan task has been used for other similar studies as it is open-ended, accepts different viewpoints, requires a short time, and does not require previous knowledge [1]. Creating slogans is representative of collaborative practices in many industries such as businesses, sports, and politics [16, 25, 34].

3.3 Chat Interface and CA Design

As shown in Figure 1, we implemented a simple text-based chat interface using HTML/CSS, Node.js, MongoDB database, and Socket.io [5]. Participants can choose their usernames to enter the chat, and they can only communicate publicly. A CA named FacilitatorBot was deployed in the chat room and was introduced using a neutral framing [4]. When the CA sent a private message, the message only appeared to the recipient and was tagged with "(PRIVATE message to @username)" upfront as demonstrated in Figure 1. We initially designed a separate chat room to deliver private messages, but we found that many participants from pilot studies did not notice private messages until the end of the study. Thus, we presented private messages in the same group chat room to control the time of interventions.

The CA monitored the participation level of individuals by calculating the standard deviation of two features, the number of messages and the number of unique words for each member, every six minutes. These two features have been used in prior works as proxies for quantity and quality of contributions [32]. The member with the lowest z-score in either feature was identified as the under-contributing member. If multiple members had the same z-score, the system randomly selected a member. This detection method allowed us to control the number of interventions across treatment conditions. Averages (standard deviations in parentheses) of the number of messages and unique words per interval were 5.32 (2.28) messages and 36.81 (17.96) words, and the values were similar across conditions ($p = 0.88$). The average number of messages and unique words for the selected under-contributing member in a group were 3.22 messages and 20.29 words, which are about 39-45% less than the average values of groups.

We tailored the messages depending on the feature: the CA asked for more elaboration when the under-contributing member had the lowest number of unique words (e.g., "Hey @username, your messages were relatively short. (..) Could you explain more about your opinion or suggest new thoughts?") or asked for general opinions when the member had the lowest number of messages (e.g., "Hey @username, you have written relatively fewer messages. (..) Would you tell your team your opinions?"). The message contents were designed based on a prior work showing that people have more motivation to participate when their contribution is identifiable, important, and has a specific goal [40]. We designed messages to be consistent across conditions as much as possible to reduce potential confounding effects. For example, we specified recipient (e.g., @username, @everyone) in both private and public conditions. We used usernames rather than real names to prevent any privacy concerns of participants and mitigate name bias.

In the peer-mediated strategy condition, the CA selected the peer who is contributing the most at the given duration because they are more likely to attend to CA's messages, to have higher social influence [43], and may even be prevented from dominating the conversation. Referring to previous studies about techniques to improve the peer's adherence [11], we provided a message template that they could refer to. The CA asks the peer privately because a CA asking a peer to prompt another person when the person can read the CA's message felt awkward by the pilot study participants.

3.4 Participants

We recruited Amazon Mechanical Turk (MTurk) workers who were at least 18 years old, located in the US, native English speakers, and often use online chats. These eligibility criteria were selected based on prior studies to mitigate time zone differences [56], proficiency of the language [23], and technology readiness among group members. Additionally, workers whose number of approved tasks was at least 1000 were allowed to work on our task; this criterion was suggested by the MTurk community[1] to receive quality responses. We asked participants to use a desktop or laptop computer for the study.

Among 962 participants who signed up, we invited 604 participants to the task who were interested in creative tasks and can join the study on time. 192 participants showed up and passed

[1]https://www.reddit.com/r/mturk/

Table 2. Demographic profiles of the participants. The average number of participants per group who are associated with each factor are reported in the Group column (standard deviations in parentheses).

Factors	Range	Total	Group Mean (SD)	Factors	Range	Total	Group Mean (SD)
Gender	Male	99	2.4 (0.72)	**Ethnicity**	White	140	3.3 (1.24)
	Female	85	2.0 (0.64)		Asian or Pacific Islander	23	0.5 (0.67)
	Prefer not to say	1	0.0 (0.15)		Black or African American	13	0.3 (0.47)
					Hispanic or Latino	8	0.2 (0.45)
					Other	1	0.02 (0.15)
Age	18-29 years	41	1.0 (0.95)	**Education**	Less than high school degree	1	0.0 (0.15)
	30-39 years	72	1.7 (1.09)		High school degree or equivalent	22	0.5 (0.74)
	40-49 years	36	0.9 (0.84)		Some college but no degree	28	0.7 (0.72)
	50-59 years	19	0.5 (0.63)		Associates degree	18	0.4 (0.59)
	60 years or older	17	0.4 (0.59)		Bachelors degree	96	2.3 (1.02)
					Graduate degree	20	0.5 (0.55)
Personality	Extroverted	86	2.0 (0.85)	**Residence**[3]	Population larger than 50,000	33	0.8 (1.02)
	Introverted	99	2.4 (0.76)		Suburb or small city	152	3.6 (1.41)

the pre-task activity, resulting in an attrition rate similar to that in prior research [26]. Eight participants dropped out during the task, mostly at the beginning of the discussion thus limiting the impact on the group dynamics. Consequently, 185 participants completed the discussion and the post-discussion survey.

The participants were distributed in 42 groups, which is seven groups per condition. We balanced gender (i.e., proportion of females in a group) and personality [22] (i.e., proportion of extroverted members in a group) across conditions using covariate adaptive randomization [30] when assigning participants to groups. We restricted that group sizes to between 3 to 6 members and they were comparable across conditions (M=4.40, SD=1.01)[2]. No significant differences were found across conditions in participants' chat usage frequency nor familiarity with CAs in general. Demographic profiles of the participants are listed in Table 2. Past research found that information diversity (e.g., major, occupation) is most closely tied with the group's performance in making creative solutions [65]. We found information diversity among participants in terms of occupation and major as participants reported approximately 110 types of occupation and 59 types of major.

3.5 Study Procedure

The user study was an online study and took about 50 minutes. We ran the study in batches of six groups. Each batch started the task at the same time in separate group chat rooms. We streamlined the process with MTurk platform based on prior works [1, 64] and suggestions from the MTurk community, which we summarized as the following:

(1) *Sign-up form*: The participants signed an online consent form, reported their demographic information and their interests in doing a creative task.
(2) *Enter waiting room*: When the scheduled time was near, people who were invited after the filtering process entered the virtual waiting room on our website.
(3) *Task information*: At the scheduled time, participants were directed to a page where they read information about the task and the compensation.
(4) *Group chat*: Participants were randomly assigned to a group chat room where they entered using a username of their choice.

[2] We used M for mean and SD for standard deviation as abbreviations
[3] https://simplemaps.com/data/us-cities

(5) *Pre-task activity*: The CA first invited participants to do a quick getting-acquainted exercise (e.g., chatting about hobbies) for about 7 minutes before the task. If a participant did not respond to the exercise, we assumed that the participant was not paying attention to the chat and removed them from the chat room before the group task started.
(6) *Group task*: Participants had 30 minutes to decide on a final slogan as a group. The CA structured the discussion based on the Diamond of Participation framework [29], which fosters divergence of ideas followed by convergence. We provided a survey link only after the timer has ended.
(7) *Post-discussion survey*: After the discussion, participants individually finished a survey that took about 10 minutes. We asked them to review the chat history during the survey to assist their memory. In the private channel conditions, the system revealed all the private messages sent by the CA after the discussion had ended so that members who did not receive any private message can get a full picture of what the CA has done for the group.

Participants were compensated $0.5 for the sign-up form and $8 for completing the task. We provided a bonus payment ($1/person) to 10% of the total number of groups based on the task outcome. We emphasized before the discussion that individual participation in the discussion is not relevant to compensation and bonuses in which all compensations will be equally distributed to members. The reason is that we did not want to give them the impression that they need to actively participate to get paid and we wanted to have more variability in participation.

3.6 Measures

We collected chat logs, survey responses, and group decision outcomes to measure participation, perceptions, and group performance.

3.6.1 Participation (RQ1, RQ4). We used the Response Quality Index (RQI) [69] to evaluate the quality of each message. RQI indicates whether members are making meaningful participation during the discussion. The RQI is calculated by multiplying ratings of three metrics – relevance, clarity, and informativeness – with each metric was rated on a 3-point scale [69]. First, three researchers labeled a small subset of the chat data and compared the results. After discussing the discrepancies and creating a coding guideline, the researchers independently labeled a test sample of the data (approx. 10%) to calculate inter-rater reliability. Krippendorff's alpha was 0.902, indicating reliable agreement [35]. The rest of the data was labeled independently by these researchers using the established guidelines.

As a proxy measure to evaluate the under-contributing members' participation, we calculated the *Gini coefficient* for the number of messages sent by each member weighted by the RQI. The Gini coefficient ranges from zero to one and indicates a degree of inequality. If a member is under-contributing significantly compared to the rest of the group, the Gini coefficient will be closer to one. The coefficient is often used in similar studies to measure participation balance [32, 52, 57].

However, it is possible to obtain high participation balance due to low overall participation; an extreme example is a situation where a perfect balance is achieved as all members sent no message. Therefore, we also checked that the *Total Participation*, which is the sum of weighted messages normalized by the group size, was similar across conditions.

3.6.2 Perceptions (RQ2, RQ4). In the post-task survey, we asked participants about *perceived effectiveness* (i.e., How effective do you think these messages are to balance participation during the discussion) and *social influence* of the CA [43] (i.e., The FacilitatorBot was influential/directive in promoting participation during the discussion) (2 items, α=0.74). We also asked perceived *intrusiveness* [39] (3 items, α=0.89) and *embarrassment* [44] (4 items, α=0.87) when receiving a message from the CA. All survey responses were recorded as seven-point Likert scales.

The survey included open-ended questions investigating the perceived advantages and disadvantages of their CA's communication strategy. Researchers individually reviewed the responses and used qualitative content analysis [19] to extract the themes. Authors iteratively discussed the themes until consensus was reached and measured the frequency of each theme. If the frequency was too low (less than 4; peer-mediated: less than 2), the theme was excluded from the list.

3.6.3 Decision performance (RQ3, RQ4). To measure the *group performance* based on the task outcomes, three researchers individually rated each slogan on a five-point scale for the following criteria: 1) how unique, unusual, or novel is this slogan, 2) how useful is this slogan for the intended purpose [1, 48]. We calculated the average rating for each slogan [48].

3.6.4 Learning effect (RQ4). We asked the following question in both the sign-up survey and the post-task survey: "Suppose you see a member who is contributing less to the discussion compared to other members in the group. How likely are you to involve that person in the discussion if no one else is taking an action?". By calculating the difference between pre- and post-task survey results, we wanted to check whether there was any learning effect in the peer-mediated strategy condition.

3.7 Statistical Analysis

We used Linear Mixed Models (LMMs) for hierarchical modeling. We used the *lme4* R package for building LMMs and the restricted maximum likelihood estimation. We constructed LMMs with experimental factors (e.g., publicness, specificity) as fixed effect variables for the full factorial analysis test. When we compared two strategies (e.g., peer-mediated vs. public individual), we used the strategy type as a fixed effect variable. For the analysis of individual survey responses, we used Group ID as a random effect variable to account for intraclass correlation (i.e., between-cluster variance to the total variance controlled by random intercept models [50]). Pertaining to group-level dependent variables (e.g., group performance), we used Batch ID as a random effect variable. In all analyses, we added the group size as a covariate (fixed) to control its effect [7]. To report statistical significance, we conducted likelihood-ratio chi-squared tests of the full model with the effect in question against the model without the effect in question (R processes can be found in [66]), which is a common approach for LMMs [45].

4 RESULTS

In this section, we explain how the publicness and the specificity of communication strategies from a CA affect participation (RQ1), perceptions about the CA (RQ2), and group performance (RQ3). We then report the results of how the peer-mediation strategy compares with the public individual strategy (RQ4). As we previously explained, we separated the peer-mediated strategy from the other results as it does not fit in the factorial analysis test for RQ1-3.

4.1 Participation (RQ1)

We examined participation balance within a group using the Gini coefficient for the number of messages weighted by the RQI; the higher the Gini coefficient, the more imbalance of participation. We conducted a LMM regression by including the Gini coefficient as dependent variable, publicness, specificity, or their interactions as fixed-effects factors, the group size as a covariate, and Batch ID as a random-effects factor (analysis explained in section 3.7). In Appendix A.1, we report detailed

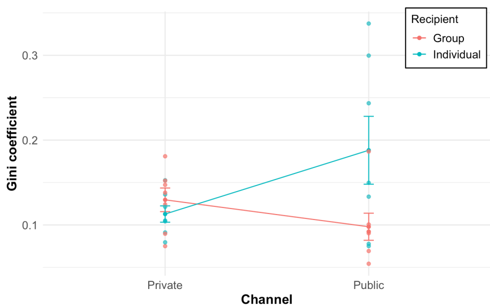

Fig. 2. There was a significant interaction effect of publicness and specificity when predicting the balance of participation. Note that the lower coefficient indicates more balanced participation.

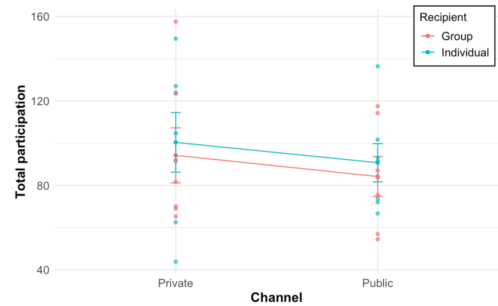

Fig. 3. The Total Participation was not significantly different across publicness and specificity.

outputs of the omnibus test result. The results showed that there was a significant interaction effect between publicness and specificity of the communication strategy ($\chi^2(1, N = 28) = 6.21, p < .05$) as shown in Figure 2. We conducted a simple main effect analysis with Bonferroni corrections to interpret the interaction effect. As a result, we found that the private individual messages (M=0.11, SD=0.03) resulted in a significantly more balanced participation compared to the public individual messages (M=0.19, SD=0.11) ($\chi^2(1, N = 28) = 6.62, p < .05$). We also found that the participation level in the public group condition (M=0.10, SD=0.04) was significantly more balanced compared to participation in the public individual condition (M=0.19, SD=0.11) ($\chi^2(1, N = 28) = 9.05, p < .01$). We did not find statistical significance in other pairs of communication strategies. For instance, the balance of participation was similar between the public group condition (M=0.10, SD=0.04) and the private group condition (M=0.13, SD=0.04) (p=.25), which is reasonable because both strategies are almost the same except the presence of the private tag (i.e., (PRIVATE message to @username)) at the beginning of the message. We conducted the same LMM regression with the Total Participation as dependent variable, and found that the Total Participation was not significantly different between publicness, specificity, or their interactions (p=.71) as depicted in Figure 3.

As an additional analysis, we calculated the differences of participation measures six minutes before and after each intervention and averaged across interventions (i.e., delta). We conducted LMM regressions with the same factors used in the previous analysis and the delta as a dependent variable. The results were consistent with our previous results. For the delta Gini coefficient, we found that there is an interaction effect of the publicness and the specificity ($\chi^2(1, N = 28) = 5.64, p < .05$) with the same pattern of simple main effects. The delta Total participation were not statistically different ($p = .19$).

We conducted planned comparisons between the proposed and the control strategies to get additional insights on our agent's effectiveness. We conducted the same LMM regressions but replaced the fixed-effects factor with the strategy type. We found the Gini coefficient of the private individual strategy was significantly lower than the control condition (M=0.15, SD=0.05), meaning more balanced participation ($\chi^2(1, N = 14) = 4.69, p < .05$). The Gini coefficient of the public group strategy was also significantly lower than the control condition ($\chi^2(1, N = 14) = 9.86, p < .01$). The Total Participation was not significantly different between the control condition and the private individual strategy (p=0.19) nor the public group strategy (p=0.73).

To summarize, we found that the *private individual strategy* and the *public group strategy* were more effective to balance participation without losing total participation of all members, compared

Table 3. Descriptive statistics of the perception measures. Participants rated the measures using seven-point Likert scales. We reported the means and standard deviations in the parentheses. Values in bold showed marginal or significant improvement compared to the values in counterparts (private vs. public, individual vs. group, peer-mediated vs. public individual).

Perceptions	Individual		Group		Control	Peer mediated
	Public	Private	Public	Private		
Perceived effectiveness	4.44 (1.54)	5.19 (1.57)	4.23 (1.80)	4.27 (1.87)	4.32 (1.68)	**5.14 (1.51)**
Intrusiveness	3.75 (1.94)	**2.89 (1.83)**	4.06 (2.05)	**3.76 (1.85)**	4.34 (1.86)	**2.86 (1.84)**
Embarrassment	3.91 (1.58)	**3.38 (1.41)**	3.58 (1.47)	**3.23 (1.67)**	4.12 (1.41)	**2.93 (1.56)**
Social influence	4.95 (1.62)	5.20 (1.65)	4.89 (1.75)	4.53 (1.61)	4.60 (1.64)	5.12 (1.79), Peer: **5.95 (0.99)**

to the public individual messages or the control condition, thus more effective in encouraging participation of under-contributing members.

4.2 Perceptions of the Agent (RQ2)

We first compare the quantitative results and then summarize the qualitative results based on the open-ended responses from the survey.

4.2.1 Quantitative findings. We asked about the perceived effectiveness, intrusiveness, and embarrassment caused by the CA's communication strategies in the post-task survey. For each perception measures, we conducted a LMM regression analysis by including the publicness, the specificity, or their interactions as fixed-effects factors, the group size as a covariate, and Group ID as a random-effects factor. For the embarrassment scale, we included a covariate of whether the member was targeted by the agent or not. Descriptive statistics about the perception measures are summarized in Table 3.

Messages sent through the private channel were perceived to be less intrusive with marginal significance[4] ($\chi^2(1, N = 125) = 3.39, p < .10$) and significantly less embarrassing ($\chi^2(1, N = 125) = 3.96, p < .05$) than messages sent through the public channel. We report detailed model outputs in Appendix A.2. However, specificity did not significantly affect perceived intrusiveness (p=0.17) nor embarrassment (p=0.44). Besides, we found no significant difference across the publicness and the specificity on the perception of social influence (p=.66) or the perceived effectiveness (p=0.11).

We conducted planned comparisons with the proposed and the control strategies to get additional insights about how the agent is perceived by users. We conducted the same LMM regressions as before but replaced the fixed-effects factor with the strategy type. Results show that the perceived effectiveness of the private individual strategy was higher than the control condition with marginal significance ($\chi^2(1, N = 63) = 3.33, p < .10$). The private individual strategy was also significantly less intrusive ($\chi^2(1, N = 63) = 5.62, p < .05$) and significantly less embarrassing than the control condition ($\chi^2(1, N = 63) = 5.45, p < .05$). Despite the improvement, it is important to note that the raw values of perceived intrusiveness and embarrassment of the agent is not zero (refer to Table 3), which is a trade-off of having agent interventions for effective teamwork. The results are still meaningful because our goal is to minimize the negative side effects of the agent by manipulating the communication strategies while increasing their effectiveness. We expect that the side effects can be further reduced by providing a better framing and adjusting the frequency of interventions.

4.2.2 Qualitative findings. Looking at Table 4 summarizing the advantages and disadvantages of each strategy, we can observe perceptions that were brought up specifically to each strategy, but

[4]We considered p<.05 as significant and p<0.10 as marginally significant, following the common practice for small-scale experiments [13]

Table 4. Advantages and disadvantages of communication strategies in each condition experienced by the participants. The number of participants who mentioned the theme is written in parentheses.

Strategy	Advantages	Disadvantages
Public Individual	Feel pressured to respond (9) Recipients of the messages are clear (9) Increased transparency (6) Encourage all members to participate more (4) Others can nudge under-contributing members (4)	Feel embarrassed and withdrawn (20) Recipients feel singled out, and even attacked (10)
Private Individual	Prevents public embarrassment (24) Does not disrupt ongoing conversations (6) Recipients of the messages are clear (5)	Easy to ignore (12) Recipients feel singled out (9) Lack of transparency (8) Others cannot nudge under-contributing members (4)
Public Group	No one feels singled out (22) Increased transparency (20) Encourages all members to participate more (7) Feel pressured to respond (5)	Recipients are unclear (17) Interrupts conversation flow (9) Easy to ignore (6) Active members get confused (5) Feel embarrassed and withdrawn (4)
Private Group	Prevents public embarrassment (10) No one feels singled out (8) Recipients feel less pressured (4)	Easy to ignore (15) Messages feel less personalized (6) Active members get confused (4) Recipients are unclear (4)
Peer Mediated	Inclusiveness (7) Natural conversation with people (4) Conversations become more friendly (2)	Makes an awkward conversation (6) Members may become annoyed towards the peer (5) Peer may ignore the message (4) Additional work on the peer (2)

we can also observe perceptions that are common across the factors. Many participants reported that they prefer private messages because it prevented public embarrassment compared to public messages: "*the bot communicated privately with me and sent a small message without embarrassing me [P29, Private group]*". When private messages were combined with the individual strategy, participants felt they are less disrupting the discussion because it is only visible to one person compared to other strategies: "*It doesn't create as much of a distraction to others who may be concentrating on brainstorming ideas [P102, Private individual]*".

On the other hand, we discovered that public messages can let other active members to see what is going on (i.e., increased transparency) and thereby creating a group accountability to engage with under-contributing members more as well as themselves: "*This (public messages) encouraged everyone to seek feedback and participation from the person who was addressed for not participating [P91, Public individual]*". While we did not see significant quantitative results, we observed that public messages successfully motivated under-contributing members to participate more and private messages can be easily ignored, which is in line with the persuasion and social influence theory [67]: "*It is harder for the person to ignore the FacilitatorBot since everyone has seen that the message has been sent [P3, Public individual]*". However, we found that public messages may not always lead to increased participation overall since it could be "*embarrassing the person, causing the person to become more withdrawn [P33, Public individual]*". In other words, while the under-contributing member briefly participated in response to the CA's public message, it may not lead to other voluntary participation when the CA was not sending the prompts.

Common opinions about the specificity were also found. One of the common advantages of the individual strategies was that "*It can better target those individuals that need to participate more, rather than leaving everybody ambiguous as to who it might be that needs to participate more [P105, Private individual]*". On the contrary, one of the advantages of the group strategy was that it did not highlight any individuals but still nudge the under-contributing members to contribute: "*A*

Table 5. Descriptive statistics of the decision performance and the learning effect for all conditions. We reported the means and standard deviations in the parentheses. There were no significant differences across conditions in the performance. We observed a positive learning effect in the peer-mediated condition.

Measures	Individual		Group		Control	Peer mediated
	Public	Private	Public	Private		
Decision performance	3.02(1.10)	2.95(0.96)	2.90(1.01)	3.12(0.95)	3.57(0.94)	3.05(0.91)
Learning effect	-0.47(1.34)	-0.06(1.39)	-0.23(1.52)	-0.17(1.21)	-0.06(1.29)	**0.24(1.27)**

person wouldn't feel singled out, but still understand that they need to do what is directed at the entire group [P44, Private group]". However, we found that group strategy can also make everyone feel they were prompted, which could possibly lead to side-effects such as overwhelming participation of all members: "*(I felt) like the Bot didn't address person/people it should have. Made it feel like we all were slacking. We weren't. [P30, Public group]*"

In sum, *private messages* resulted in lower perceived intrusiveness and embarrassment than public messages. Qualitative responses show that there are *trade-offs* for every strategy. For example, the private individual strategy can prevent public embarrassment, reduce distraction, and help people to easily identify who the under-contributing members are. However, the strategy can be easily ignored and the recipient can feel singled out.

4.3 Decision Performance (RQ3)

We examined whether the communication strategies had an impact on task outcomes. We conducted a LMM regression by including the averaged slogan ratings as dependent variable, the publicness, specificity, or their interactions as fixed-effects factors, the group size as a covariate, and Batch ID as a random-effects factor. We saw no significant difference in slogan ratings for the publicness and the specificity ($p = 0.97$) nor for the planned comparisons with the control condition ($p = 0.79$). We initially hypothesized that more voices participating will lead to diversity of ideas and produce a more creative outcome [15] but it was not supported in our study. Dijk et al. argued that the impact of diversity on group performance can be small, especially for less complicated tasks [59]. Therefore, it is possible that our slogan making task was too easy and short to gain a significant difference in performance. The descriptive statistics are summarized in Table 5.

4.4 Peer-mediated Strategy (RQ4)

In this section, we compare the peer-mediated strategy with the public individual strategy. As we previously explained in Section 3.1, both strategies are in common how the under-contributing members are prompted, thus make a fair comparison. For each participation-related scale (i.e., Gini coefficient, Total participation, delta), we conducted LMM regressions by including the experimental condition (peer-mediated vs. public individual) as a fixed-effects factor, the group size as a covariate, and Batch ID as a random-effects factor. Similar analysis was done for perception-related scales with Group ID as a random-effects factor. In Appendix A.3, we reported detailed statistical results of the effects. We only report the patterns of interest. As we summarized the descriptive statistics in Table 3, we found that the peer-mediated strategy can improve how users perceive the agent. Participants perceived that the agent using the peer-mediated strategy was more effective to balance participation with marginal significance ($\chi^2(1, N = 61) = 3.63, p < .10$), significantly less intrusive ($\chi^2(1, N = 61) = 4.29, p < .05$), and significantly less embarrassing ($\chi^2(1, N = 61) = 7.19, p < .01$) compared to the public individual strategy. Another interesting finding was that the participants felt that the peer's social influence (i.e., being influential or directive) is significantly higher than the agent's social influence in the public individual condition ($\chi^2(1, N = 61) = 5.31, p < .05$).

When we asked about whether they are likely to involve under-contributing members in discussions before and after the study, we found a higher learning effect with marginal significance in the peer-mediated condition compared to the public individual condition ($\chi^2(1, N = 61) = 3.72, p < .10$). As summarized in Table 5, there was a positive learning effect in the peer-mediated condition (M = 0.24, SD = 1.27), meaning that participants were more likely to involve the under-contributing members in the discussion after the study, in contrast to a negative learning effect we found for the public individual condition (M = -0.47 SD = 1.34). The control condition showed almost no difference in learning effect before and after the task (M = -0.06, SD = 1.29). This is interesting because an agent with the peer-mediation design helped people to learn how to engage with others, whereas the conventional design can have an opposite effect. It could be that participants receiving the public individual strategy believed the CA would intervene instead of them, thus resulted in more reliance on the CA and less motivation to involve others in discussions.

We additionally analyzed the open-ended responses about their perceptions of the peer-mediated strategy. As summarized in Table 4, participants reported inclusiveness as the main strength of the peer-mediation approach. For example, participants mentioned that the CA's messages can promote interactions among members (e.g., "*It (CA) gets people more directly involved with each other, allowing the bot to only facilitate rather than become involved itself [P127, Peer-mediated]*"), under-contributing members felt included (e.g., "*It's nice to know your peers want to hear what you have to say. It can also make your peers be more mindful of making sure you're included if you struggle to speak up moving forward [P128, Peer-mediated]*"), and peers felt they were important (e.g., "*It makes me feel like a leader of the group as I was asked to encourage others to speak. It makes me feel as if my input was important and useful [P134, Peer-mediated]*").

Another main advantage of the peer-mediated strategy was that the peers crafted personal and contextual messages that make the conversation flow more natural. For example, instead of using a generic message template that the CA suggested (e.g., could you tell us your opinion?), many participants contextualized the message such as "*@Kathy, what's your favorite at this point?*". However, some participants reported that it may feel awkward to be selected as the mediator and to have a conversation with the under-contributing member: "*It felt weird to be singled out as I don't usually see myself as a leader and it seemed like the bot was asking me to be one [P152, Peer-mediated]*". We expect that the awkwardness can be reduced by members taking turns to perform the mediator's role or asking for volunteers prior to the chat.

To summarize, we found that the *peer-mediated strategy* has potentials to improve the perceptions of the agent compared to the conventional strategy. While the differences in the participation and performance were not significant, we observed a learning effect that people are more likely to involve under-contributing members in the discussion.

5 DISCUSSION

We found that the **private individual strategy** is the most effective strategy to promote participation of under-contributing members, decrease the perceived embarrassment and intrusiveness of the CA, compared to other public or group strategies. We also offered a preliminary evidence that the peer-mediated strategy benefits how users perceive the agent, reducing perceived embarrassment and intrusiveness of the CA, compared to the conventional approach where the CA directly prompts an under-contributing member. In the following paragraphs, we discuss design implications, followed by future research directions and limitations.

5.1 Design Implications for Facilitator CAs

The results showed that the public messages specifying an individual recipient was not effective in balancing participation. This result contradicted our initial expectation based on prior research that under-contributing members are more likely to respond to the message when they are publicly and individually prompted due to higher normative pressures [67] and more social influence from the CA [43] compared to other strategies. From qualitative findings, we discovered that while participants might have felt pressured to respond to public individual messages, some retreated from the discussion afterward due to the unsolicited attention they got. We hypothesize that this short-lived participation could have risen from different causes of under-participation. For participants who under-contribute due to the lack of motivation or unawareness of their low participation, a public call out could motivate them to increase their participation throughout the session. However, under-participation could also be based on low confidence or evaluation apprehension [3]. For these participants, the public individual strategy could only elicit temporal participation rather than a sustained engagement. Therefore, it could be important to identify the causes of under-contribution to determine which strategy to use.

The peer-mediated strategy helped people to learn how to engage with others. We expect that the strategy has a positive impact on group members, even in spaces where CAs are not utilized, by giving them learning opportunities about various roles they can play for the group. However, we found that the peer-mediated strategy did not balance the participation more than the public individual strategy, partly because some peers ignored the agent's request. Prior research mentioned that people prefer not to address the under-contributing members due to relational burden [32]. From survey responses, we further discovered that some peers ignored the CA's messages when 1) they were focusing on the discussion rather than performing the mediator's role, 2) the person to address had just participated, or 3) their prior intervention did not get a response. These observations lead to design implications that when using the peer-mediated strategy, the CA should choose a peer who is not too deeply engaged at the time of intervention; otherwise the CA's request can be overlooked. Next, our CA offered a generic message for the peers to use as a template, but presenting more context-relevant message templates would help the peer relay a more appropriate message. Third, it is important that a different peer is selected for consecutive interventions to distribute the responsibility to multiple peers and reduce their burden.

Every strategy has trade-offs, thus integrating multiple types of interventions within a single CA can be useful. For example, a CA may initially address under-contributing members using the private individual strategy. However, if one person has been selected too many times, the CA can change to a public group strategy to not stress the person and bring other members' attention. If the overall participation needs to be high rather than a particular person's participation, the group strategy could be used so that everyone can increase awareness of their participation. If an inclusive atmosphere and the learning experience during the discussion are important such as in educational settings, the peer-mediated strategy is recommended. In consideration of these design implications, future research can be considered to iteratively design facilitator CAs with rich qualitative and trace data.

5.2 Agent Design in Group Discussions

We considered various aspects when designing the facilitator CA such as how to achieve balanced participation and how to configure various parameters regarding group characteristics and contexts. First, we designed a CA that promotes participation of under-contributing members to achieve balanced participation. We used a text-based chat for discussions which can hold multiple conversational floors so that one's dominating behavior is less likely to take away the opportunity of others [55]. However, there are circumstances where over-participating members block the voices of other participants, such as video conferences where only one conversational floor is allowed at a time. In such cases, a CA that requests dominating members to yield their turn can be studied to further balance participation in groups.

The ideal distribution of participation might differ based on the group size, duration, and task. For example, for large groups, it can be simply impractical to expect equal participation within a short meeting. When having multiple discussions over a long period of time, CA should go beyond looking at the balance of participation within a single discussion to measuring individual contributions from a long-term perspective. If one person comes up with the correct answer to a problem-solving task where there is only one solution, it may not be important to have additional opinions from other members. As such, CA designers should be aware of the ideal distribution of participation they are designing for depending on various group dimensions.

While we recruited native English speakers from MTurk for a task that does not require prior knowledge, it should be noted that the design of a facilitator CA may vary depending on the characteristics of each member. For example, individuals' linguistic skills, knowledge levels and familiarity on the task, and their attitudes towards AI can have a significant impact on their participation. It is desirable for a CA to first assess individuals' linguistic and task-relevant skills and give assistance on alleviating language barriers or accessing necessary information when needed, rather than blindly urging participation. For people who are generally averse to AI technologies, passive methods such as intervening only when requested can be used to interact with them. It is also important to consider the effect of existing relationships between members. For example, in a group with hierarchical relationships, low-status members may find it harder to have a voice, thus a CA empowering their voices while asking high-status members to listen to opinions from low-status members can be useful.

Our suggested communication strategies could serve as a starting point for CAs in other domains. The suggested design could be explored in online classrooms such as an agent privately messaging a student to contribute to the discussion. Additionally, our findings could offer insights into the design of moderator bots for online governance. For example, an agent that detects misinformation in an online community should effectively communicate with the poster as well as other members before removing the post. The peer-mediated strategy could be beneficial to other discussions where the feeling of inclusiveness is valued such as peer support group chats. Follow-up studies can investigate other dimensions and modalities of interventions such as message tones, presentation methods (e.g., pop-up windows), questions or informative styles, explicit or implicit prompts, as well as other variations of peer-mediated messages.

5.3 Limitations

First, this work only addressed a particular task and a group setting, thus more work is needed to test the generalizability for other types of tasks, group sizes, and group dynamics. Our results may also be limited to the recruited demographics (e.g., MTurk workers). Second, we had only seven groups per condition which may have limited the statistical power. Third, we found lack of effects on the decision outcome. Future research should explore complex tasks that may benefit more from opinions of diverse people. Fourth, the CA used a simple detection method to select under-contributing members. Due to its simplicity, it is possible that the CA may have falsely identified an engaged person as under-contributing, especially those who made invisible contributions (e.g., people who are actively thinking) or short high-quality responses. Using better heuristics or a more accurate ML-based detection algorithm in future studies will resolve the issue.

6 CONCLUSION

In this paper, we evaluated five communication strategies that a CA could employ to address an under-contributing member. Our results showed that the private individual strategy is the most effective and preferred strategy overall by improving the balance of participation, while reducing the perceived intrusiveness from the agent and embarrassment. Additionally, we found that the peer-mediated strategy can enhance perceptions of the agent and offer learning opportunities about how to engage other members, compared to the strategy where a CA intervenes directly. Qualitative findings further identified that each strategy has trade-offs and the preferred strategy could change depending on the group dynamic and context. Our research findings and discussions about the strengths and weaknesses of each strategy can inform the design of the facilitator agent for group discussions. Based on our findings that the communication strategy can have a significant influence on the effectiveness and perception of a CA, we encourage future research on different type of communication strategies that account for various group characteristics and contexts.

REFERENCES

[1] Faez Ahmed, Nischal Reddy Chandra, Mark Fuge, and Steven Dow. 2019. Structuring Online Dyads: Explanations Improve Creativity, Chats Lead to Convergence. In *Proceedings of the 2019 on Creativity and Cognition*. 306–318.

[2] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N Bennett, Kori Inkpen, et al. 2019. Guidelines for human-AI interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems*. 1–13.

[3] Yair Amichai-Hamburger, Tali Gazit, Judit Bar-Ilan, Oren Perez, Noa Aharony, Jenny Bronstein, and Talia Sarah Dyne. 2016. Psychological factors behind the lack of participation in online discussions. *Computers in Human Behavior* 55 (2016), 268–277.

[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] Damien Arrachequesne. 2021. Socket.io Chat Tutorial. https://socket.io/demos/chat/

[6] Anthony D Bergstrom. 2011. *Social mirrors: visualization as conversation feedback*. Ph.D. Dissertation. University of Illinois at Urbana-Champaign.

[7] Markus Brauer and John J Curtin. 2018. Linear mixed-effects models and the analysis of nonindependent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychological Methods* 23, 3 (2018), 389.

[8] Julia Cambre, Scott R Klemmer, and Chinmay Kulkarni. 2017. Escaping the echo chamber: ideologically and geographically diverse discussions about politics. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 2423–2428.

[9] Julia Cambre, Ying Liu, Rebecca E Taylor, and Chinmay Kulkarni. 2019. Vitro: Designing a Voice Assistant for the Scientific Lab Workplace. In *Proceedings of the 2019 on Designing Interactive Systems Conference*. 1531–1542.

[10] Joel Chan, Steven Dang, and Steven P Dow. 2016. Improving crowd innovation with expert facilitation. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. 1223–1235.

[11] Jeffrey M Chan, Russell Lang, Mandy Rispoli, Mark O'Reilly, Jeff Sigafoos, and Heather Cole. 2009. Use of peer-mediated interventions in the treatment of autism spectrum disorders: A systematic review. *Research in autism spectrum disorders* 3, 4 (2009), 876–889.

[12] Eshwar Chandrasekharan, Chaitrali Gandhi, Matthew Wortley Mustelier, and Eric Gilbert. 2019. Crossmod: A cross-community learning-based system to assist reddit moderators. *Proceedings of the ACM on human-computer interaction* 3, CSCW (2019), 1–30.

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

[14] Gayle V Davidson-Shivers, Lin Y Muilenburg, and Erica J Tanner. 2001. How do students participate in synchronous and asynchronous online discussions? *Journal of Educational Computing Research* 25, 4 (2001), 351–366.

[15] Carsten KW De Dreu and Michael A West. 2001. Minority dissent and team innovation: The importance of participation in decision making. *Journal of applied Psychology* 86, 6 (2001), 1191.

[16] Dawn R Deeter-Schmelz, Karen Norman Kennedy, and Rosemary P Ramsey. 2002. Enriching our understanding of student team effectiveness. *Journal of Marketing Education* 24, 2 (2002), 114–124.

[17] Steven Dow, Julie Fortuna, Dan Schwartz, Beth Altringer, Daniel Schwartz, and Scott Klemmer. 2011. Prototyping dynamics: sharing multiple designs improves exploration, group rapport, and results. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 2807–2816.

[18] Gregory Dyke, David Adamson, Iris Howley, and Carolyn Penstein Rosé. 2013. Enhancing scientific reasoning and discussion with conversational agents. *IEEE Transactions on Learning Technologies* 6, 3 (2013), 240–247.

[19] Satu Elo and Helvi Kyngäs. 2008. The qualitative content analysis process. *Journal of advanced nursing* 62, 1 (2008), 107–115.

[20] David Engel, Anita Williams Woolley, Ishani Aggarwal, Christopher F Chabris, Masamichi Takahashi, Keiichi Nemoto, Carolin Kaiser, Young Ji Kim, and Thomas W Malone. 2015. Collective intelligence in computer-mediated collaboration emerges in different contexts and cultures. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 3769–3778.

[21] Thomas Erickson, David N Smith, Wendy A Kellogg, Mark Laff, John T Richards, and Erin Bradner. 1999. Socially translucent systems: social proxies, persistent conversation, and the design of "babble". In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 72–79.

[22] Samuel D Gosling, Peter J Rentfrow, and William B Swann Jr. 2003. A very brief measure of the Big-Five personality domains. *Journal of Research in personality* 37, 6 (2003), 504–528.

[23] Zixuan Guo and Tomoo Inoue. 2019. Using a conversational agent to facilitate non-native speaker's active participation in conversation. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–6.

[24] F Maxwell Harper, Dan Frankowski, Sara Drenner, Yuqing Ren, Sara Kiesler, Loren Terveen, Robert Kraut, and John Riedl. 2007. Talk amongst yourselves: inviting users to participate in online conversations. In *Proceedings of the 12th international conference on Intelligent user interfaces*. 62–71.

[25] Rune Høigaard, Filip Boen, Bert De Cuyper, and Derek M Peters. 2013. Team identification reduces social loafing and promotes social laboring in cycling. *Int J Appl Sports Sci* 25, 1 (2013), 33–40.

[26] James Houghton. 2019. Launching multiplayer games (poor click-through?). https://forum.turkerview.com/threads/launching-multiplayer-games-poor-click-through.2419/#post-1280158

[27] Jeroen Janssen, Gijsbert Erkens, Gellof Kanselaar, and Jos Jaspers. 2007. Visualization of participation: Does it contribute to successful computer-supported collaborative learning? *Computers & Education* 49, 4 (2007), 1037–1065.

[28] Malte F Jung, Nikolas Martelaro, and Pamela J Hinds. 2015. Using robots to moderate team conflict: the case of repairing violations. In *Proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction*. 229–236.

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

[30] Minsoo Kang, Brian G Ragan, and Jae-Hyeon Park. 2008. Issues in outcomes research: an overview of randomization techniques for clinical trials. *Journal of athletic training* 43, 2 (2008), 215–221.

[31] Sara Kiesler and Lee Sproull. 1992. Group decision making and communication technology. *Organizational behavior and human decision processes* 52, 1 (1992), 96–123.

[32] Soomin Kim, Jinsu Eun, Changhoon Oh, Bongwon Suh, and Joonhwan Lee. 2020. Bot in the Bunch: Facilitating Group Chat Discussion by Improving Efficiency and Participation with a Chatbot. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[33] 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*. 457–466.

[34] Chiranjeev Kohli, Lance Leuthesser, and Rajneesh Suri. 2007. Got slogan? Guidelines for creating effective slogans. *Business horizons* 50, 5 (2007), 415–422.

[35] Klaus Krippendorff. 2004. Reliability in content analysis: Some common misconceptions and recommendations. *Human communication research* 30, 3 (2004), 411–433.

[36] Bibb Latané, Kipling Williams, and Stephen Harkins. 1979. Many hands make light the work: The causes and consequences of social loafing. *Journal of personality and social psychology* 37, 6 (1979), 822.

[37] Sung-Chul Lee, Jaeyoon Song, Eun-Young Ko, Seongho Park, Jihee Kim, and Juho Kim. 2020. SolutionChat: Real-time Moderator Support for Chat-based Structured Discussion. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–12.

[38] 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.

[39] Hairong Li, Steven M Edwards, and Joo-Hyun Lee. 2002. Measuring the intrusiveness of advertisements: Scale development and validation. *Journal of advertising* 31, 2 (2002), 37–47.

[40] Kimberly Ling, Gerard Beenen, Pamela Ludford, Xiaoqing Wang, Klarissa Chang, Xin Li, Dan Cosley, Dan Frankowski, Loren Terveen, Al Mamunur Rashid, et al. 2005. Using social psychology to motivate contributions to online communities. *Journal of Computer-Mediated Communication* 10, 4 (2005), 00–00.

[41] Ioanna Lykourentzou, Shannon Wang, Robert E Kraut, and Steven P Dow. 2016. Team dating: A self-organized team formation strategy for collaborative crowdsourcing. In *Proceedings of the 2016 CHI conference extended abstracts on human factors in computing systems*. 1243–1249.

[42] Lotte Meteyard and Robert AI Davies. 2020. Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language* 112 (2020), 104092.

[43] Michael D Miller and C Cryss Brunner. 2008. Social impact in technologically-mediated communication: An examination of online influence. *Computers in Human Behavior* 24, 6 (2008), 2972–2991.

[44] Andre Modigliani. 1971. Embarrassment, facework, and eye contact: Testing a theory of embarrassment. *Journal of Personality and social Psychology* 17, 1 (1971), 15.

[45] Christopher H Morrell. 1998. Likelihood ratio testing of variance components in the linear mixed-effects model using restricted maximum likelihood. *Biometrics* (1998), 1560–1568.

[46] Clifford Nass, Jonathan Steuer, and Ellen R Tauber. 1994. Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 72–78.

[47] 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 & Virtual Environments* 12, 5 (2003), 481–494.

[48] Michelle A Pang and Carolyn C Seepersad. 2016. Crowdsourcing the evaluation of design concepts with empathic priming. In *ASME 2016 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. American Society of Mechanical Engineers Digital Collection.

[49] Sherry L Piezon and William D Ferree. 2007. Perceptions of social loafing in online learning groups. In *23 rd Annual Conference on Distance Teaching & Learning, http://www. uwex. edu.*

[50] José Pinheiro and Douglas Bates. 2006. *Mixed-effects models in S and S-PLUS*. Springer Science & Business Media.

[51] Alisha Pradhan, Leah Findlater, and Amanda Lazar. 2019. "Phantom Friend" or "Just a Box with Information": Personification and Ontological Categorization of Smart Speaker-based Voice Assistants by Older Adults. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–21.

[52] Gianluca Schiavo, Alessandro Cappelletti, Eleonora Mencarini, Oliviero Stock, and Massimo Zancanaro. 2014. Overt or subtle? Supporting group conversations with automatically targeted directives. In *Proceedings of the 19th international conference on Intelligent User Interfaces*. 225–234.

[53] Kay Kyeongju Seo. 2007. Utilizing peer moderating in online discussions: Addressing the controversy between teacher moderation and nonmoderation. *The American Journal of Distance Education* 21, 1 (2007), 21–36.

[54] Ameneh Shamekhi, Q Vera Liao, Dakuo Wang, Rachel KE Bellamy, and Thomas Erickson. 2018. Face Value? Exploring the effects of embodiment for a group facilitation agent. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–13.

[55] James Simpson. 2005. Conversational floors in synchronous text-based CMC discourse. *Discourse studies* 7, 3 (2005), 337–361.

[56] John C Tang, Chen Zhao, Xiang Cao, and Kori Inkpen. 2011. Your time zone or mine? A study of globally time zone-shifted collaboration. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work (CSCW)*. 235–244.

[57] 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.

[58] Stergios Tegos, Stavros Demetriadis, and Thrasyvoulos Tsiatsos. 2014. A configurable conversational agent to trigger students' productive dialogue: a pilot study in the CALL domain. *International Journal of Artificial Intelligence in Education* 24, 1 (2014), 62–91.

[59] Hans Van Dijk, Marloes L Van Engen, and Daan Van Knippenberg. 2012. Defying conventional wisdom: A meta-analytical examination of the differences between demographic and job-related diversity relationships with performance. *Organizational Behavior and Human Decision Processes* 119, 1 (2012), 38–53.

[60] Fernanda B Viegas and Judith S Donath. 1999. Chat circles. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 9–16.

[61] Erin Walker. 2010. *Automated adaptive support for peer tutoring*. Ph.D. Dissertation. Carnegie Mellon University.

[62] Erin Walker, Nikol Rummel, and Kenneth Koedinger. 2009. Beyond explicit feedback: new directions in adaptive collaborative learning support. *In Proceedings of the Computer Supported Collaborative Learning Practices*, 552–556.

[63] Qiaosi Wang, Koustuv Saha, Eric Gregori, David Joyner, and Ashok Goel. 2021. Towards Mutual Theory of Mind in Human-AI Interaction: How Language Reflects What Students Perceive About a Virtual Teaching Assistant. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.

[64] Xu Wang, Miaomiao Wen, and Carolyn Rosé. 2017. Contrasting explicit and implicit support for transactive exchange in team oriented project based learning. Philadelphia, PA: International Society of the Learning Sciences.

[65] Sam R Wilson, William C Barley, Luisa Ruge-Jones, and Marshall Scott Poole. 2020. Tacking Amid Tensions: Using Oscillation to Enable Creativity in Diverse Teams. *The Journal of Applied Behavioral Science* (2020), 0021886320960245.

[66] Bodo Winter. 2013. Linear models and linear mixed effects models in R with linguistic applications. *arXiv:1308.5499. [http://arxiv.org/pdf/1308.5499.pdf]* (2013).

[67] Wendy Wood. 2000. Attitude change: Persuasion and social influence. *Annual review of psychology* 51 (2000).

[68] Ziang Xiao, Michelle X Zhou, Wenxi Chen, Huahai Yang, and Changyan Chi. 2020. If I Hear You Correctly: Building and Evaluating Interview Chatbots with Active Listening Skills. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.

[69] Ziang Xiao, Michelle X Zhou, Q Vera Liao, Gloria Mark, Changyan Chi, Wenxi Chen, and Huahai Yang. 2020. Tell Me About Yourself: Using an AI-Powered Chatbot to Conduct Conversational Surveys with Open-ended Questions. *ACM Transactions on Computer-Human Interaction (TOCHI)* 27, 3 (2020), 1–37.

[70] Amy X Zhang and Justin Cranshaw. 2018. Making sense of group chat through collaborative tagging and summarization. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–27.

[71] Roshanak Zilouchian Moghaddam, Zane Nicholson, and Brian P Bailey. 2015. Procid: Bridging consensus building theory with the practice of distributed design discussions. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. 686–699.

7 APPENDIX

A MODEL AND TEST OUTPUTS

We summarize full model outputs, following the best practice guideline [42]. We only report the models of major findings that showed statistical significance or marginal significance (p-values highlighted in bold). Note that the p−values of each fixed effect were estimated via t−tests using the Satterthwaite's method, which is a common method of reporting LMM models. We flag levels of significance ($^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$). We rounded all values to two decimal places.

A.1 Participation (RQ1)

Table 6. Participation balance (Gini coefficient)

Likelihood ratio test						
Effect		logLik		χ^2		p
Publicness+Specificity+Publicness:Specificity		41.65		8.98		**0.03****
Fixed Effects						
	Estimate (β)	SE	95% CI		t	p
Intercept	0.07	0.05	-0.03	0.16	1.36	0.19
Group size	0.01	0.01	-0.01	0.04	1.45	0.16
Publicness	-0.03	0.03	-0.09	0.03	-1.16	0.26
Specificity	-0.02	0.03	-0.08	0.04	-0.71	0.48
Publicness:Specificity	0.11	0.04	0.03	0.19	2.64	0.01**
Random Effects						
				Variance		S.D.
Batch ID (Intercept)				2.08e-22		1.44e-11
Residual				2.989e-03		5.467e-02
Model fit						
R^2				Marginal		Conditional
				0.33		0.33

A.2 Perceptions (RQ2)

Table 7. Perceived intrusiveness

Likelihood ratio test						
Effect		logLik		χ^2		p
Publicness		-254		3.39		**0.07***
Fixed Effects						
	Estimate (β)	SE	95% CI		t	p
Intercept	5.62	0.80	4.04	7.19	7.04	1.1e-10***
Group size	-0.44	0.16	-0.77	-0.12	-2.70	0.01***
Publicness	0.61	0.33	-0.04	1.26	1.85	0.07*
Specificity	-0.46	0.33	-1.11	0.20	-1.37	0.17
Random Effects						
				Variance		S.D.
Group ID (Intercept)				0.00		0.00
Residual				3.39		1.84
Model fit						
R^2				Marginal		Conditional
				0.1		0.1

Table 8. Perceived embarrassment

Likelihood ratio test						
Effect		logLik		χ^2		p
Publicness		-225		3.96		**0.05****
Fixed Effects						
	Estimate (β)	SE	95% CI		t	p
Intercept	3.12	0.67	1.80	4.43	4.66	7.9e-06***
Group size	-0.08	0.13	-0.34	0.18	-0.60	0.55
Target member	0.71	0.27	0.18	1.24	2.63	0.01***
Publicness	0.53	0.27	0.01	1.06	2.01	0.05**
Specificity	0.21	0.27	-0.32	0.73	0.78	0.44
Random Effects						
				Variance		S.D.
Group ID (Intercept)				0.00		0.00
Residual				2.15		1.47
Model fit						
R^2				Marginal		Conditional
				0.08		0.08

A.3 Peer-mediated Strategy (RQ4)

Table 9. Perceived effectiveness

Likelihood ratio test						
Effect		logLik		χ^2		p
Condition		-111		3.63		**0.06***
Fixed Effects						
	Estimate (β)	SE	95% CI		t	p
Intercept	3.67	0.89	1.89	5.45	4.11	0.00***
Group size	0.16	0.18	-0.19	0.51	0.90	0.37
Condition	0.75	0.39	-0.02	1.51	1.93	0.06*
Random Effects						
					Variance	S.D.
Group ID (Intercept)					0.00	0.00
Residual					2.22	1.49
Model fit						
R^2					Marginal	Conditional
					0.06	0.06

Table 10. Perceived intrusiveness

Likelihood ratio test						
Effect		logLik		χ^2		p
Condition		-123		4.29		**0.04***
Fixed Effects						
	Estimate (β)	SE	95% CI		t	p
Intercept	5.52	1.09	3.35	7.69	5.06	4.1e-06***
Group size	-0.36	0.21	-0.79	0.06	-1.70	0.10*
Condition	-1.0	0.47	-1.93	-0.06	-2.1	0.04**
Random Effects						
					Variance	S.D.
Group ID (Intercept)					0.00	0.00
Residual					3.31	1.82
Model fit						
R^2					Marginal	Conditional
					0.10	0.10

Table 11. Perceived embarrassment

Likelihood ratio test						
Effect		logLik		χ^2		p
Condition		-111		7.19		**0.01*****
Fixed Effects						
	Estimate (β)	SE	95% CI		t	p
Intercept	5.41	0.96	3.50	7.31	5.66	0.00***
Group size	-0.33	0.18	-0.68	0.02	-1.87	0.07
Target member	0.23	0.39	-0.54	1.00	0.59	0.55
Condition	-1.07	0.39	-1.84	-0.30	-2.76	0.01***
Random Effects						
					Variance	S.D.
Group ID (Intercept)					0.00	0.00
Residual					2.22	1.49
Model fit						
R^2					Marginal	Conditional
					0.15	0.15

Table 12. Perceived social influence

Likelihood ratio test						
Effect		logLik		χ^2		p
Condition		-100		5.31		**0.02***
Fixed Effects						
	Estimate (β)	SE	95% CI		t	p
Intercept	3.64	0.99	1.59	5.73	3.67	0.00***
Group size	0.26	0.20	-0.16	0.68	1.32	0.21
Condition	1.12	0.45	0.19	2.06	2.51	0.02**
Random Effects						
					Variance	S.D.
Group ID (Intercept)					0.38	0.62
Residual					1.29	1.13
Model fit						
R^2					Marginal	Conditional
					0.18	0.37

Table 13. Learning effect

Likelihood ratio test						
Effect		logLik		χ^2		p
Condition		-101		3.72		**0.05***
Fixed Effects						
	Estimate (β)	SE	95% CI		t	p
Intercept	0.74	0.76	-0.76	2.24	0.98	0.33
Group size	-0.25	0.15	-0.54	0.05	-1.68	0.10*
Condition	0.64	0.33	-0.01	1.29	1.96	0.06*
Random Effects						
					Variance	S.D.
Group ID (Intercept)					0.00	0.00
Residual					1.59	1.26
Model fit						
R^2					Marginal	Conditional
					0.11	0.11

Received July 2021; revised November 2021; accepted April 2022