Bot in the Bunch: Facilitating Group Chat Discussion by Improving Efficiency and Participation with a Chatbot

Soomin Kim[1] Jinsu Eun[1] Changhoon Oh[2] Bongwon Suh[1] Joonhwan Lee[1]
[1]Seoul National University, Seoul, Korea [2] Carnegie Mellon University, Pittsburgh, USA
[1]{soominkim, eunjs71, bongwon, joonhwan}@snu.ac.kr [2]ochangho@andrew.cmu.edu

ABSTRACT
Although group chat discussions are prevalent in daily life, they have a number of limitations. When discussing in a group chat, reaching a consensus often takes time, members contribute unevenly to the discussion, and messages are unorganized. Hence, we aimed to explore the feasibility of a facilitator chatbot agent to improve group chat discussions. We conducted a needfinding survey to identify key features for a facilitator chatbot. We then implemented GroupfeedBot, a chatbot agent that could facilitate group discussions by managing the discussion time, encouraging members to participate evenly, and organizing members' opinions. To evaluate GroupfeedBot, we performed preliminary user studies that varied for diverse tasks and different group sizes. We found that the group with GroupfeedBot appeared to exhibit more diversity in opinions even though there were no differences in output quality and message quantity. On the other hand, GroupfeedBot promoted members' even participation and effective communication for the medium-sized group.

Author Keywords
Chatbot; Conversational agent; Group chat; Discussion; Consensus; Online communication; GroupfeedBot.

CCS Concepts
•**Human-centered computing** → **Human computer interaction (HCI);** *User studies;*

INTRODUCTION
Group chat is pervasive in everyday life and used not only for social but also for transactional purposes [17, 35]. For example, group chat serves as an effective channel for decision-making, problem-solving, and open debates on specific issues [35, 61]. These goal-oriented communications are common in diverse groups, including those of family members, friends, colleagues, and coworkers. Group chat has the advantages of enabling asynchronous communication [39, 52] and maintaining users' awareness of ongoing group agendas through traceable chat logs [42]. These features enable comfortable group communication among members in various locations and with various temporal schedules [8] without causing disruption in their daily lives [33].

Despite these advantages, a number of difficulties are common when conducting goal-oriented discussions through a group chat. First, reaching timely consensus can be more difficult in a group chat than in a face-to-face meeting, as procrastination and loss of concentration are common in a group chat [25]. Second, messenger-mediated interaction without in-person contact can lead to uneven participation, which reduces the satisfaction of those who do actively participate and thus weakens positive group dynamics [20, 30]. This problem is exacerbated for larger groups because it is hard to even detect that a member has become a "lurker"—someone who does not actively participate and plays a passive role in group interactions [1, 38]. Third, group chat's unstructured, unthreaded chat interface causes difficulty in the organization of diverse opinions. Human moderators thus often seek to restructure and/or summarize fragmented messages; however, this requires high cognitive load, lowering discussion efficiency [42].

Research on group discussions has focused on supporting work-related groups [12, 24, 42, 46] by structuring [4, 61], summarizing [14, 49], and visualizing [23, 27] opinions. This study aims to advance this line of research by focusing on social groups. Furthermore, we extend the previous work by supporting group discussions with a novel strategy of facilitating even participation among group members.

In this paper, we aim to design and develop a chatbot agent that can enhance group discussions in social chat groups. We propose GroupfeedBot, a chatbot that can act as a facilitator in group discussions by (1) managing time, (2) encouraging members to participate evenly, and (3) organizing the members' diverse opinions. It should be noted that the focus of this study is on goal-oriented communication (e.g., decision-making, problem-solving, and open discussion) rather than on socially oriented communication (e.g., chitchat, joke-making, and talk about daily life).

To explore the feasibility of a chatbot agent in group chat discussions, we executed a series of user studies. Through a needfinding survey, we define a number of the features that a facilitating chatbot agent should possess. We then implemented GroupfeedBot, a chatbot agent meant to support the group chat discussions. Figure 1 shows how GroupfeedBot

CHI '20, April 25–30, 2020, Honolulu, HI, USA.

ACM ISBN 978-1-4503-6708-0/20/04 ...$15.00.
DOI: http://dx.doi.org/10.1145/3313831.3376785

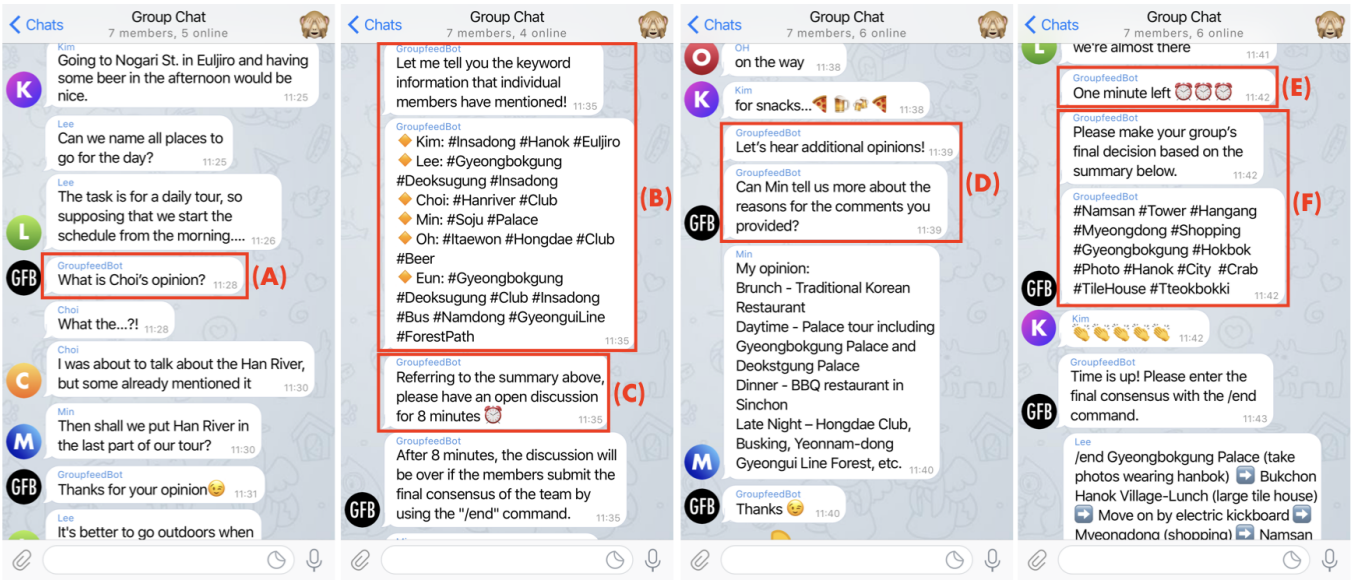

Figure 1. Conversational design strategies applied in GroupfeedBot to facilitate group chat discussions. The chatbot manages the discussion time (1C, 1E), facilitates even participation by encouraging lurkers to speak up (1A, 1D), and organizes individual members' (1B) and overall groups' (1F) opinions. All messages are translated from Korean.

facilitates discussions in group chats. This facilitator chatbot is meant to support discussions in group chats by managing time (1C, 1E), encouraging lurkers to speak up (1A, 1D), organizing individual members' opinions (1F), and summarizing overall opinions (1B).

To verify the effectiveness of GroupfeedBot, we conducted a qualitative study with small-sized groups followed by a user study with medium-sized groups. The qualitative study involved six small groups, each composed of four to five members (N = 25). We conducted in-depth semi-structured interviews to gain deeper insights into the chatbot agent's features.

We then conducted the preliminary user study with two medium-sized groups of 10 members each (N = 20). A mixed factorial design with one between-subjects variable (type of chatbot: basic vs. GroupfeedBot) and one within-subjects variable (type of task: estimation vs. decision-making vs. social debating vs. problem-solving) was used. We measured group behavior (message quantity, opinion diversity, and even participation), users' attitudes (communication efficiency/effectiveness/openness, and usefulness), and output quality. We also conducted post-hoc focus group interviews (FGI). The results indicate the following:

- There were no difference in message quantity and output quality between the groups that discussed with GroupfeedBot and those that discussed with the basic chatbot. However, the group which discussed with GroupfeedBot tended to produce more diverse opinions.
- GroupfeedBot encouraged the members to contribute evenly to the discussions, especially for the open-debating task. Furthermore, the members perceived the group chat sessions as involving fairer and more effective communication.
- Some effects of GroupfeedBot varied by the task's characteristics. It led to high communication efficiency and perceived usefulness in the decision-making and open-debating task but not in the estimation and problem-solving tasks.

This paper makes the following contributions:

- We verified that a chatbot agent can be effective not just in dyadic but also in group interactions.
- We designed and developed a chatbot agent to improve discussion in group chat by enabling timely consensus, facilitating even participation, and organizing opinions.

RELATED WORK

Improving goal-oriented communication in distributed teams has been a focus of research in computer-supported cooperative work (CSCW) since the field's inception [34, 56]. This study on the use of chatbots in group discussions can be viewed with respect to multiple research areas, including mobile instant messaging (MIM) and group chats, online group discussions, and chatbots' applications in group communication.

Mobile Instant Messaging and Group Chat

The accompanying proliferation of MIM has changed the ways in which people interact. In particular, as compared to SMS, one of the distinguishing features of MIM is that its group chat allows for asynchronous multiparty communication [7]. Because group chat is convenient for communicating with multiple members, groups with various types of relationships use it for a variety of purposes (e.g., chitchat, information exchange, and cooperative works) [7, 9, 50].

Group chat researchers have focused on the ways in which such chats can improve collaboration in the workplace [12, 24, 42, 46]. However, given the widespread use of MIM, group

chats have become common in general as well; even groups with informal relationships have cooperative team processes. To reflect these trends, researchers have carried out studies to improve the generally perceived inconveniences [55, 61]. Along the same lines as those researchers, we sought to resolve the unexpected inconveniences of using group MIM chats.

Systems for Supporting Group Discussions

Previous studies in the area of online group discussion have explored various ways of supporting collaboration among remote users of various collaborative systems, including email, online documents, and internet relay chat. The most studied method for improving online discussion is summarization. In particular, scholars have shown that summarization can improve email systems [14, 49] by preventing information overload [11]. Moreover, researchers have effectively applied social annotations and other feedback to discussions involving online documents or dashboards [6, 48, 62]. Another line of research on support for group collaboration is focused on improving group awareness with visualizing diverse members' point of view [23, 27]. These and other methods use visualization components to decrease the cognitive effort needed to synthesize diverse members' opinions [22].

Unlike threaded forums or dashboards, chat interfaces lack a reply-based interface; thus, researchers have sought to support group discussion by embedding more structure in chats. For instance, groups in a text chat discussion are more likely to reach consensus when preauthorized structures and scripts are provided [15]. Such a script structure allows for more focused discussion by providing procedural guidance. OpinioNetIt automatically organizes members' opinions on controversial topics, in terms of support or opposition [4]. Similarly, Tilda, the Slack bot, is meant to enhance the group sense-making process by organizing chat messages using features such as tagging, linking, and summarization [61].

In light of the findings from previous studies, we consider a new way of driving active participation among users and of decreasing their mental loads. We focused on improving discussions by involving the communication "process" and participation patterns. Furthermore, the chat interface used in MIM with its speech balloons requires an approach for improving discussions that is distinct from those that are used for emails, dashboards, and forums. Few researchers have explored the use of text-based virtual agents that can perform specific functions that are targeted to the MIM system. Our work aims to understand users' needs when doing discussion in group chat and to design a chatbot agent.

Chatbots in Group Interactions

Chatbots are widely used for the dyadic interactions in various fields, including user research [36, 57], customer support [28, 60], and health care [37]. Most previous works in these areas involve dyadic chatbots that interact directly with users. However, researchers have made several attempts to examine chatbots' effects in the context of group interaction.

Multiparty-based chatbots are virtual agents that communicate with multiple users in a group. Researchers on multiparty-based chatbots have regarded chatbots as tools that support group interaction. For instance, Calendar.Help [10] enables efficient time scheduling for groups through structured workflows. In addition, TaskBot [58] mediates task management in a chatroom through functions such as assigning tasks, terminating tasks, and setting reminders. Furthermore, SearchBots [2] assists with collaborative information-seeking by presenting users with the information that they have requested.

Although researchers who have studied chatbots in group interaction have tended to consider chatbots in terms of their support for goal-oriented tasks, some scholars have recently extended this research to include chatbots as group members. Those researchers have focused on chatbots' social roles. In the Reddit community, bots serve functions such as administrating, providing play or humor, ensuring functionality or quality, fostering community, and archiving [40]. Seering et al. [53] identified the role of bots in the Twitch community by analyzing bot messages. In subsequent work, Seering et al. [54] proposed seven new social roles for chatbots, drawing from a research-through-design approach.

The results of these studies raise important questions about the feasibility of chatbots acting as community members. In this work, we utilized a community-based perspective which views chatbots to be group members rather than tools.

NEEDFINDING SURVEY FOR FACILITATOR CHATBOT AGENT

A needfinding survey was performed to discover the features that a facilitator chatbot should possess.

Method and Participants

We recruited an initial sample of 134 participants from Amazon Mechanical Turk. We asked them open-ended questions about the context of the group discussion, the users' positive and negative experiences with group chats, and their desired improvements to group chats. To filter out unqualified responses, we omitted 31 participants for incorrect responses to the filtering item ("Please enter the name of the mobile instant messenger you used for the group chat."). We also excluded 43 participants whose median response times were less than 240 seconds and whose responses were not sincere. Accordingly, we finally obtained 65 valid responses. The final respondents had a mean age of 30.83 years (SD = 7.96; 29 female).

Analysis

To discover themes in the data, we applied thematic analysis based on the bottom-up approach to the open-ended responses. Three researchers coded the open-ended answers, and the process was repeated five times until the identified themes were saturated. These final codes were then organized within the main themes from which we derived the implications.

Findings and Design Goals for Facilitator Chatbot

Based on the survey results, we categorized three main findings and created three design goals for the facilitating chatbot agent ($\kappa = 0.72$).

Quick Communication Leads to Efficient Discussion.
We noted that time management was important to efficient discussion. Several respondents mentioned that quick responses

create a well-interacting group: "Well-interacting group members are quick and efficient and do not procrastinate" (P65); "Quick and efficient communication and response times are crucial" (P23). On the other hand, unrestricted discussion caused dissatisfaction: "The worst meeting is just out of control and goes on at all hours of the day and night" (P8). In addition, certain members' late responses hindered efficient discussion: "People who do not answer promptly ruin the discussion" (P55). Members who did not respond in a timely manner not only caused inefficiencies but also worsened the group's mood. For example, P44 noted, "They don't text back often and don't answer until days or weeks later. They make the group feel insecure and depressed."

- Design Goal 1: Support the ability to reach consensus in a timely manner by ensuring an efficient discussion procedure.

Even Contributions Enhance Group Coherence.
The members' engagement and participation is an important factor in both individual members' satisfaction and overall group coherence. A well-functioning discussion is characterized by even contributions from all members. P36 mentioned, "In a well-interacting group, everyone says something." Similarly, P18 stated, "Good discussion is made when everyone contributes evenly. Everybody gets along, the conversations are nonstop, everyone gets an opinion, no one bashes each other." However, lurkers can appear to be representative members actually interrupt positive discussions. P14 noted, "Lurkers silently take in all that is happening in the group but do not post. These people do not contribute." Lurkers lower a group's coherence and undermine the satisfaction of the active members; participants complained about "stand-backers who don't really care to be in the group" (P20) and "apathetics who do not really care much about the group" (P47).

Other types of members can prevent positive interactions as well; for instance, the "chatterbox" type can also prevent constructive and fair discussion, as P28 noted: "Gabbers seem to have nothing better to do than to talk all day and night. They often overshare and dominate conversations with talk of their personal problems." P76 concurred, saying that chatterboxes "hog the convo, are dismissive of some members, and ignore other members. These hoggers make most members feel left out." To summarize, even participation is critical to a group discussion's communication quality, as well as to members' satisfaction and positive group dynamics.

- Design Goal 2: Encourage even contributions to ensure coherence and satisfaction among the group members.

Organizing Diverse Members' Opinions Is Challenging.
We discovered that the process of organizing and synthesizing diverse opinions was important in ensuring high-quality discussion outputs. Each team needs an organizer or leader to ensure good discussions: "There is the organizer, someone who can make sense of the whole discussion and wrap things up" (P8); "I usually organize the group chat and keep things in order" (P49). However, these organizers could not participate sufficiently in the discussion because synthesizing and organizing the members' opinions required a lot of effort [61, 51]. Many participants expected the chatbot to make the discussion process smoother by reconstructing the opinions of various group members: "A chatbot could be useful to help organize the group chat and sort the conversation into threads for easy reading" (P33); "I expect that the chatbot will clean up duplicated messages and organize messages better" (P2).

- Design Goal 3: Aggregate and organize diverse opinions to reduce participants' cognitive load.

GROUPFEEDBOT: A CHATBOT AGENT FOR FACILITATING DISCUSSION IN GROUP CHATS

Based on the findings from the survey, we designed GroupfeedBot. This chatbot runs on the Telegram messaging application and was built with BotFather. The back-end server was built with Python, using the Telegram library and pickleDB. The front-end and back-end use a Telegram dispatcher to communicate, transmit data, and access APIs. GroupfeedBot was designed and developed with the following features:

Time Management

GroupfeedBot sets up time limitations for each task to efficiently derive consensus within a restricted time (Figure 1C). Time setting encourages group members to be on the same page and to collaborate in achieving the discussion goals. It also uses a time alert to induce the members to organize their opinions (Figure 1E).

Encourage Lurkers to Speak Up

To encourage even contributions, GroupfeedBot engages members who have not spoken. GroupfeedBot detects the members who had not commented and asks, "What is *[the lurker name]*'s opinion?" (Figure 1A). As the second strategy for driving even participation, the chatbot asks the less outspoken members for additional comments. It counts the number of words that each member has submitted and then asks the person with the lowest total to elaborate on their opinions with the question, "Can *[the lurker name]* tell us more about the reasons for the comments you provided?" (Figure 1D).

GroupfeedBot gave a proper message based on lurker's response. When GroupfeedBot asked for a lurker's opinion, if the lurker responded within 90 seconds, GroupfeedBot said "Thanks for your opinion" and then moved on to the next step. Otherwise, GroupfeedBot said "Please tell me later."

Organizing Individual Group Members' Opinion

GroupfeedBot organizes the main opinions of each member (Figure 1B). It uses the Text-rank algorithm [43] to summarize each member's comments in one or two sentences. Subsequently, it extracts the lexical morphemes from the summarized sentences and presents them in the form of hashtags. This function is designed to reduce the difficulty of capturing individual members' opinions which is caused by the chat interface's unthreaded presentation method.

Summarizing Overall Opinion

GroupfeedBot summarizes comments from the entire discussion (Figure 1F). It uses the Text-rank algorithm [43] to summarize the entire team's output in four to five sentences; it then

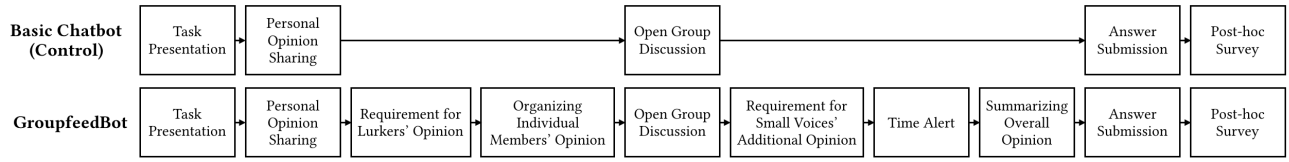

Figure 2. Task flow of the two chatbots used in the experiment. In the Groupfeedbot condition, four features to facilitate group discussions were applied.

extracts the key morphemes from the summarized sentences and presents them as hashtags.

QUALITATIVE STUDY WITH SMALL-SIZED GROUP

We performed two user studies to examine GroupfeedBot's feasibility in group chats with diverse discussion tasks and various group sizes. In the first phase, we conducted a qualitative study with small groups.

Study Design and Procedure

We performed the qualitative study for six groups, each composed of four or five members. Each group performed four tasks: two with the basic chatbot and two with GroupfeedBot. We wanted to get in-depth feedback by allowing every user to participate in the two types of discussion equally.

We gathered each randomly assigned group in a spacious room, where we briefly described the study. We then placed each participant in a separate space to reproduce non-face-to-face situations as well as possible. We invited all the participants to join a group chat in Telegram using their smartphones. Afterward, the participants again gathered in one space, where we conducted a roughly 40-minute FGI.

Participants

We recruited participants by posting an announcement on our institution's online-community website. We recruited 25 participants (14 female). Their mean age was 28.40 (SD = 3.65). To partially control for prior experience, we required that all the participants had experience using Telegram and group chatting in MIM. The study was conducted on strangers. We thought GroupfeedBot would be more viable on nascent social groups which have flexible norms so that members are more easily influenced by GroupfeedBot's nudging.

Apparatus

The chatbot was applied in two conditions: the basic chatbot and GroupfeedBot. We implemented the aforementioned four features only for the condition with GroupfeedBot to verify the chatbot's effects by comparing the two conditions. We determined the appropriate duration through the pilot test. Users wanted a longer discussion on topics which should derive consensus from diverse opinions (decision-making and open-debating tasks) rather than topics with optimal answers (estimation and problem-solving tasks). We designed the study procedure so that participants in both conditions could perform the task in the same amount of time.

- Basic chatbot without facilitation: In this condition, participants conducted a discussion without the main features of the facilitator chatbot (Figure 2). Following the basic chatbot's instructions, the participants shared individual opinions and held open group discussions.
- GroupfeedBot with facilitation: Participants performed the task with the help of GroupfeedBot (Figure 2). GroupfeedBot asked for the lurkers' opinions and organized the individual members' opinions in the early phase of the discussion. In the middle of the open group discussion, the chatbot asked for additional comments from the members who had contributed the least. When about a quarter of the time remained, the chatbot gave a time alert and summarized the overall opinions of the group.

Task

We constructed four group discussion tasks such that each group would engage in discussion for its task and submit a final answer drawn from that discussion. We selected diverse tasks that could be completed in the natural group chat context.

- Estimation Task: The participants solved simple estimation problems that involved inferring the height of the Eiffel Tower and the calories in one avocado without searching the Internet [47].
- Decision-Making Task: We chose a travel task, which is a widely used type of group decision-making task [27, 61]. Specifically, we asked the participants to plan a one-day tour of Korea for a foreign friend who was visiting the country for the first time.
- Open-Debating Task: In this task, we asked the members to provide their opinions about a moral machine dilemma [3] so as to elicit varying opinions from the members. Open debates require the evaluation and reasoning of ethical and social issues.
- Problem-Solving Task: We asked the participants to determine how best to find out a person's name that you have forgotten without directly asking for that information; the scenario involved a social context [6].

Results

We constructed a thematic map based on the participants' responses to the FGI. Three major themes emerged ($\kappa = 0.80$).

Time management induces efficient discussion.

The first theme involved efficient consensus reaching through time management. The participants compared their experiences using messenger-mediated (with no chatbot) and chatbot-mediated group discussions. Group chat discussions tend to be

lagged in the current MIM system, but with the chatbot agent, the time restrictions made the conversation more focused: "It is very useful to just bookmark the beginning and end of a discussion" (P21). Many participants noted that the chatbot was effective at helping them to manage time accurately and objectively; in short, it acted like a moderator. P12 mentioned, "Usually, when there is a meeting through mobile messenger, everyone tends not to concentrate, but the chatbot managed the discussion time so that members could immerse themselves, which makes for efficient discussion." Participants felt that GroupfeedBot's time-alert feature accelerated efficient decision-making: "It was good to present accurate time guidelines so that discussions were not lengthy. Reminding the remaining time helped us to round out the discussion" (P7).

Organizing opinions supports synthesizing process.
GroupfeedBot was particularly effective for the decision-making task that required the members to synthesize and organize their comments. Specifically, it helped the group to reach consensus by allowing the members to more effectively identify each other's views. P7 noted, "It's not [normally] easy to remember who said what, but it was comfortable [in this case] because the chatbot summed up the members' comments." On the other hand, the participants who discussed with the basic chatbot had difficulty synthesizing the comments, as that required a large amount of mental effort. P13 commented, "We had to decide on a course of travel, and it was bothersome to extract the common denominator from the members." This corresponds with previous findings that organizing diverse opinions improves group's decision-making process [4, 61].

In addition, to derive the final consensus, the members used the chatbot's summary as a reliable reference. P22 mentioned, "It was nice to see what ideas came up when it showed the keywords." The facilitator chatbot also reminded members of neglected issues as the conversation went on, which allowed the discussion to proceed without bias: "The overall summary helped us to revisit the matters that we'd forgotten because the conversation was going in a different direction" (P4).

Users perceive chatbots as group members.
The participants recognized the chatbot as a group member who played a specific role in the discussion. The participants described the chatbot using personal pronouns ("he" or "she") and as a "manager," "supporter," "facilitator," "emcee," "moderator," and "assistant." This implies that chatbots can be extended to serve as group members who engage in specific social and transactional roles, in addition to acting as a tool that assists in group discussion [54]. The participants mentioned that the chatbot served as a group member: "The chatbot participated in the convo and was not awkward, so it felt really like a human. The feeling of having a manager in the conversation helped us to have an efficient discussion" (P11); "The chatbot was like a moderator. There was no need for a separate moderator" (P25); and "I hope the chatbot gets more involved in the conversation" (P12).

USER STUDY WITH MEDIUM-SIZED GROUP

We performed a user study to examine whether GroupfeedBot could improve discussions for medium-sized groups. Based on previous work [35, 16], we defined the number of members in the medium-sized group as 10. We assumed that, as far as encouraging members to contribute evenly, the facilitator chatbot would be more effective in a medium-sized group than in a small group. This is because small groups tend to be active, whereas members in larger groups tend to show low participation [35]. Members of large groups are also more likely to engage in social loafing and free riding [29], so we wanted to explore the facilitator chatbot's feasibility in the larger (medium-sized) group. In addition, when more members participate in a discussion, it takes more cognitive load for them to organize and synthesize their various opinions, so the facilitator chatbot could be more effective in group chats with more members.

Study Design

This study used a 2×4 mixed factorial design, with chatbot (basic chatbot vs. GroupfeedBot) as a between-subjects variable and task (estimation vs. decision-making vs. open debating vs. problem-solving) as a within-subjects variable. Two groups of 10 participants for each group participated in the study (N = 20), one group had discussion with the basic chatbot and the other group with GroupfeedBot.

Procedure

As the same in the qualitative study, participants were gathered in a spacious room and placed in a separate place. They were invited to join a group chat in Telegram using their smartphones. The participants then completed the four discussion tasks. At the end of each task, the participants responded to a post-hoc survey composed of eight questions (on usefulness and on the efficiency, effectiveness, and fairness of communication). After completing all the tasks, the participants answered three open-ended questions about the group chat experience. Afterward, the participants again gathered in one space, where we conducted a roughly 30-minute FGI discussing their experiences.

Apparatus, Task, and Participant

The apparatus and tasks used in this study were identical to those used in the qualitative study. We recruited 20 participants (8 female; M_{age} = 26.15, SD_{age} = 2.32) who all had experience using Telegram and group chats. We randomly assigned the participants to one of the two chatbot conditions.

Measures

We used the measures in terms of (1) group behavior, (2) users' attitudes, and (3) output quality. We collected three forms of data: chat log, quantitative-survey data, and qualitative data (from the open-ended survey and the FGI). The chat log data includes the messages' contents, the times, the team ID, the sender ID, and the task ID. We used the number of messages, number of senders, and sending times to analyze group behaviors. We used the survey and FGI data to evaluate the users' attitudes. We used the contents of each team's final answers on each task to measure the discussion output quality.

Group Behavior
- Message Quantity: Message quantity was measured by the number of morphemes used within a group. A morpheme was used as the unit because in Korea, spacing is not based

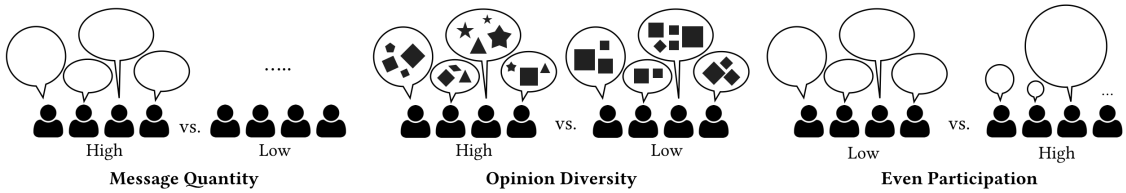

Figure 3. A graphical representation of the group behavioral variables. (A) Message quantity refers to how active the members participate in the discussion. (B) Opinion diversity is about the degree to which diversified messages are generated within the group. (C) Even participation means that how fairly and equally individual members participate in the discussion.

Measure	*Item*
Efficiency	"The chatbot helps us more *[easily or quickly]* reach a consensus as a group."
Effectiveness	"The chatbot helps us more *[confidently or comfortably]* reach consensus as a group."
Fairness	"The chatbot helps us more *[openly or fairly]* participate in the discussion"
Usefulness	"The chatbot is useful," and 'The chatbot helps me more effective."

Table 1. Survey items used for users' attitudes

on words. We could infer how actively members participated in the discussion by analyzing message quantity [19].

- Opinion Diversity: Opinion diversity is defined as the number of unique lexical morphemes shared within a group. We collected all the messages generated within a group and counted the number of unique lexical morphemes. For instance, although the lexical morpheme of "protection" was mentioned a number of times by multiple members, it was counted as one opinion unit. The message diversity is used as the approximation of the breadth of the discussion [6].

- Even Participation: We counted the number of morphemes per participant to determine how evenly the members contributed to the discussion. We used standard deviation (SD) and the Gini coefficient as the criteria for estimating even participation. The SD was standardized (divided by the time unit), so the variance of participation can be compared both by chatbot and by task. If a group has a high SD, then it has high variance in participation, which can be interpreted as indicating uneven member contributions. Similarly, the Gini index used as a measure of inequality [13]. A higher Gini index indicates greater concentration, which means a more uneven contribution.

Users' Attitudes (Quantitative and Qualitative)

After each task, the participants answered questionnaires to assess (1) communication efficiency [27], (2) communication effectiveness [27], (3) communication fairness, and (4) usefulness [41]. Each was measured with two items. The participants rated these items on a 7-point Likert scale from 1 (*strongly disagree*) to 7 (*strongly agree*). We aggregated the responses at the individual level and then at the group level. The survey items are presented in Table 1.

We also gathered qualitative responses using open-ended questions to gain more insight into the users' experience. Based on the survey results, we conducted a FGI to gain more insight into the users' experiences with the chatbot system.

Output Quality

To assess the quality of the discussion output, we evaluated the teams' final submitted answers for each task. We recruited five human judges for this task, and they rated the teams' answers on a 10-point differential scale, with higher scores indicating higher output quality. These ratings have a significant inter-rater reliability (Krippendorff's $\alpha = 0.78$).

Results

Our result has revealed that the group with GroupfeedBot marginally exchanged more diverse opinions within a group. However, there was no significant difference in the message quantity and output quality. Behavioral patterns of even participation to the discussion differed depending on the chatbot conditions. In general, participants with GroupfeedBot contributed more evenly to the discussion. Furthermore, the significant differences were observed in users' attitudes.

Group Behavior

- Message Quantity

Although the discussions had similar duration across the chatbot conditions, the mean number of words used per task was higher with GroupfeedBot ($M_{GFB} = 329.9$, $SD_{GFB} = 140.79$) than with the basic chatbot ($M_{basic} = 271.9$, $SD_{basic} = 157.53$). However, the Mann-Whitney U test revealed no significant differences between the two groups ($U = 10.00$, $Z = 0.398$, $p = 0.690$). This is may be due to the large variance in the data (Figure 4) and due to the small number of participants.

- Opinion Diversity

We conducted a Mann-Whitney U test to compare the basic chatbot and GroupfeedBot in terms of opinion diversity. An analysis of the number of unique lexical morphemes revealed that the group discussed with GroupfeedBot shared more diversified messages than did those with the basic chatbot ($U = 2.00$, $Z = 2.148$, $p = 0.032$). This result could be interpreted that GroupfeedBot's facilitating feature might increase divergent opinion exchange despite the small sample size.

- Even Participation

The overall SD and Gini index were lower in the GroupfeedBot condition than those in the basic chatbot condition. This

	All (Average)			A. Estimation			B. Decision-making			C. Open-debating			D. Problem-solving		
	Basic	GFB		Basic	GFB		Basic	GFB		Basic	GFB		Basic	GFB	
Group Behavior															
Quantity	271.9	329.9		430	526		874	1107		942	844		473	822	
Diversity	**236.3**	**318.3**	*	185	243		279	396		291	324		190	310	
Evenness (SD)	0.073	0.052		0.055	0.058		0.086	0.095		0.129	0.065		0.056	0.098	
Evenness (Gini)	0.296	0.225		0.334	0.363		0.306	0.337		0.380	0.268		0.278	0.395	
Users' Attitudes															
Efficiency	**4.41**	**5.24**	**	4.70	4.85		**3.80**	**5.35**	*	**3.55**	**5.25**	**	5.60	5.50	
Effectiveness	**3.70**	**4.75**	**	**3.60**	**3.55**	***	**3.20**	**5.20**	***	**3.65**	**5.05**	***	**4.35**	**5.20**	***
Fairness	**3.91**	**5.16**	***	**4.05**	**4.20**	***	**3.65**	**5.45**	***	**3.05**	**5.40**	***	**4.90**	**5.60**	***
Usefulness	**4.02**	**4.98**	*	4.10	4.55		**3.65**	**5.30**	**	**3.60**	**5.00**	*	4.75	5.10	
Output Quality	6.60	6.71		5.81	6.01		7.38	6.57		6.22	6.85		7.02	7.39	

Note: $^{+}p<0.10$;$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

Quantity (Message Quantity): The number of morphemes used within a group.

Diversity (Opinion Diversity): The number of unique lexical morphemes shared within a group.

Evenness (Even Contribution): The distribution of shared messages per member within a group.

Table 2. Results of the medium-sized group. In general, the group with GroupfeedBot tended to generate more diverse opinions (*Diversity*). It is noticeable that the message quantity was higher in the condition with the basic chatbot, but opinion diversity was greater with GroupfeedBot for the open-debating task. In terms of even participation, participants in the GroupfeedBot condition generally contributed to the discussion more evenly, but these effects varied depending on the task (*Evenness*). GroupfeedBot effectively led to even contribution, especially in the open-debating task. In terms of users' attitudes, GroupfeedBot participants generally perceived that the discussions were more efficient, effective, and fair, and they rated the usefulness of the chatbot agent higher. Statistically significant differences in users' attitudes are in bold.

result implies that GroupfeedBot may encourage even participation among group members. However, the even participation pattern was different depending on task type (Figure 4). GroupfeedBot has appeared to elicit even participation for the open-debating task. It may be inferred that the facilitator chatbot may induce comments from the lurkers for discussions in which personal opinion expression is of high importance.

Users' Attitude (Quantitative)

A nonparametric test of Aligned Rank Transform was conducted for mixed factorial design to test whether the main effects and interaction effect exist [59].

The Aligned Rank Transform for perceived user attitude yields a main effect for the chatbot type in terms of communication efficiency ($F(1, 18) = 10.03$, $p = 0.005$), communication fairness ($F(1, 18) = 12.33$, $p = 0.002$), communication effectiveness ($F(1, 18) = 27.47$, $p = 0.000$), and usefulness ($F(1, 18) = 10.07$, $p = 0.005$). This means that the facilitator chatbot condition produced better communication quality and usefulness than did the control condition.

We found a significant interaction between the chatbot type and the task type in terms of communication efficiency ($F(3, 54) = 7.86$, $p = 0.000$), communication fairness ($F(3, 54) = 5.81$, $p = 0.002$), communication effectiveness ($F(3, 54) = 7.73$, $p = 0.000$), and usefulness ($F(3, 54) = 3.35$, $p = 0.025$). Thus, the task type may affect the facilitator chatbot's effectiveness.

The results of the post-hoc analyses using Tukey's HSD indicate that the facilitator chatbot condition produced better communication effectiveness and communication fairness in all tasks, as compared to the control condition. In terms of communication efficiency and usefulness, the facilitator chat-

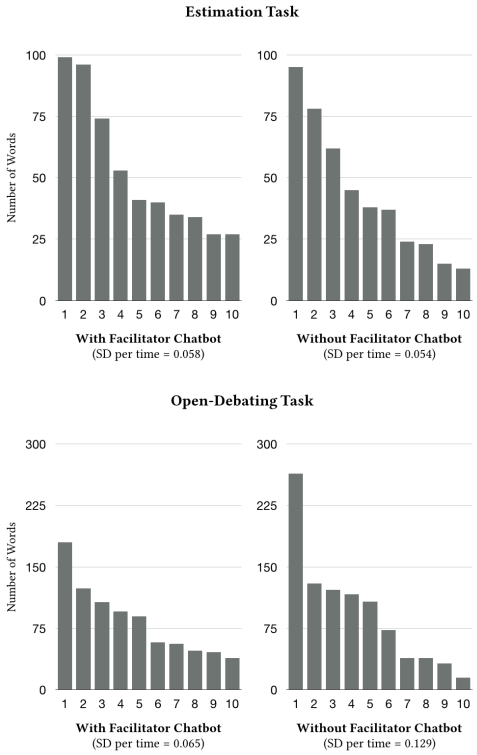

Figure 4. Distribution of messages per participant in the experiment. The even participation patterns were revealed differently depending on the tasks' characteristics. For the estimation task, there was no variation in terms of even participation by chatbot type. On the other hand, for the open-debating task, participants in the GroupfeedBot condition appeared to participate evenly in the discussion.

bot was more effective for only the decision-making (communication efficiency: $p = 0.008$; usefulness: $p = 0.000$) and open-debating (communication efficiency: $p = 0.002$; usefulness: $p = 0.019$) tasks.

These results suggest that the facilitator chatbot's effects on users' attitudes may be inconsistent across different task types. GroupfeedBot's effects were more pronounced for the decision-making and open-debating tasks. For the decision-making and the open-debating tasks, it is important to consider how many of the members spoke up and how well their diverse opinions were synthesized into the process. In these tasks, the facilitator chatbot may function well. On the other hand, for the estimation and problem-solving tasks, the chatbot's roles of facilitating diverse members' contributions and organizing their opinions were mostly irrelevant to the communication efficiency and usefulness because, in those tasks, finding the optimal answer is more important than encouraging diversity.

Member Attitude (Qualitative)

The qualitative analysis in this study focuses on the feature of encouraging even participation, the effect of which could be more pronounced in the medium-sized group ($\kappa = 0.75$).

- Encouraging Even Participation

GroupfeedBot's feature of encouraging lurkers to participate allowed the members to contribute more evenly to the discussion. Participants with the basic chatbot reported negative experiences such as having a few members dominate the discussion: "It was a shame that the opinions were skewed to one side and that the conversation was focused on a few people" (P27). On the other hand, participants in the facilitator chatbot's discussion indicated that the members contributed equally to the discussion and that the discussion went smoothly: "It was good that the chatbot distributed the opportunity to speak by pointing out who had not participated" (P39). P45, whom the chatbot called out in this manner, said, "I didn't talk much, but I was able to share my thoughts with the other members because the chatbot encouraged me to share my opinions." The involvement of various members contributed to the discussion being fair, as P40 noted: "The chatbot asked for comments from participants who didn't speak enough, which helped the team make fair decisions." P36 agreed: "Teams typically make decisions based on one member who talks a lot or who has strong opinions, but in this case, the decision was made fairer because the chatbot asked questions of the members who hadn't spoken or given many opinions."

- Reducing the Relational Burden

GroupfeedBot pointed out which members had participated less in the discussions, which reduced the relational burden that can arise when human members perform the same actions as other members. People usually feel uncomfortable calling out a fellow member; the participants in the facilitator chatbot's discussion, however, noted that they were appreciated that the chatbot replaced this kind of behavior. P41 noted, "If someone doesn't talk in a group chat, directly asking them to give their opinion is socially burdensome. It seems like you are attacking that person, so the person whom you call out can get offended. However, it was nice to have a chatbot use its neutral stance to call out members because there no misunderstanding could occur—unlike when humans talk to humans. We can't blame the bot because it is totally objective." P30 added, "It is inconvenient for a leader to nag passive members, and it would be nice for the chatbot to take one for the team instead of making the human leader do this."

- Inducing Opinion Deliberation

GroupfeeBot asks for additional comments; this led some participants to think about providing concrete evidence for their superficial claims. Deliberation can improve the quality of the participants' knowledge and opinions [21], positively affect interpersonal trust [18], and increase participation [45]. P44 noted, "I talked only a little, and the chatbot asked me for additional comments, which led me to think more deeply about the reasons for my opinion. I think it's better when the chatbot encourages additional comments more often." The facilitator chatbot's feature of encouraging additional feedback simultaneously improved both the quantity and the quality of the discussion, as P36 noted: "The chatbot asked the less talkative members for more comments, which helped to improve the quality and quantity of ideas."

Output Quality

Our analysis reveals that there were no significant differences in the output quality for any of the tasks. The discussion outputs were quite similar regardless of the chatbot's intervention style. This result suggests that, while the facilitator chatbot's involvement did not lead to negative consequences, it could increase the discussants' satisfaction and encourage even participation.

DISCUSSION

In this section, we discuss the findings of the study and its implications for designing a conversational agent for group discussions.

Supporting Group Dynamics by Facilitating Interaction in Social Groups

Our study has implications that go beyond dyadic interactions with a chatbot agent. A chatbot agent could efficiently support discussion procedures and facilitates communication, thus enhancing both taskwork and teamwork [5]. Previous research has introduced diverse methods to improve goal-oriented discussions in distributed groups [6, 23, 22, 27, 61]. Extending those studies, we explored whether a chatbot agent could support group discussions for social groups. Furthermore, we proposed a new method to improve discussions, driving members' even participation by encouraging lurkers to speak. We focused on how individual members interact within a group because the communication process is the "essence of the social system [31]." Reciprocal opinion exchange exposes members to diverse ideas and helps to improve understanding of different points of view [26], which eventually improves group dynamics. GroupfeedBot could increase positive group dynamics by encouraging all members to contribute.

Chatbot as a Group Member Not as a Tool

Our work supports the notion that chatbots are capable of serving as group members, not merely as tools [54]. Until recently,

chatbot agents were treated as tools for the use of improving work processes, particularly for goal-oriented tasks [2, 10, 58]. Recent works are more likely to refer to those agents as partners and treat them as members of a team [40, 53, 54]. This new metaphor indicates a broader and more social-oriented set of tasks, which we expect chatbots to support. Proceeding from this perspective, we designed GroupfeedBot and explored its feasibility as a facilitator of social group discussion. Previous research found that bots could serve as social members in Twitch communities, by providing diverse functions including sharing of information, explaining moderation, and promoting the streamer [53]. We sought to extend such research by exploring whether chatbots could also facilitate the group chat discussions of social groups.

Our results also imply that the chatbots' roles can evolve over time as a group dynamic evolves. For instance, GroupfeedBot called out a member who had not participated in the beginning of the discussion; that member then actively participated in the rest of the discussion. If a chatbot promotes group dynamics in this way during the early stage, then it can play additional roles after healthy group norms have been constructed; for instance, it can serve as a social organizer by pairing members or groups with similar interests or opinions [54].

Considering Group and Relationship Characteristics

Although we conducted a lab study to verify GroupfeedBot's core effects, future work could focus on diverse variables related to groups in the wild. GroupfeedBot's features were more pronounced in medium-sized groups, in which it is hard to discern every member's participation status and lurkers are more likely to occur. These results imply that group characteristics and discussion contexts should be considered when applying chatbots in the wild. Relationship type should be considered as well. From the FGI results, 9 participants noted that GroupfeedBot might be more effective in groups with a formal social affiliation (e.g., school reunion) rather than informal social groups. Group maturity could also matter. The issues of how GroupfeedBot affects the formation and adjustment of group norms and how to design a chatbot that supports groups over a longer period of time can be interesting future venues.

Simple Messages Can Nudge User Participation

Notably, the chatbot's simple nudge of asking members for their opinions induced actual behavior. When the members were prompted to add to the discussion, they shared their points of view with the group more often. In the medium-sized group study, GroupfeedBot identified six lurkers, five of whom responded to GroupfeedBot's request. Individual members do not engage in social loafing and negligence simply because they are performing tasks in a group. Rather, social loafing occurs because an individual's contribution/non-contribution to the group is not identifiable [32]. In messenger-mediated situations where social loafing is difficult to detect, the facilitator chatbot could drive member participation by making the lurkers' undesirable behaviors identifiable.

Chatbot's Effects Can Vary Depending on Tasks

In terms of even participation and users' attitudes, the facilitator chatbot's effectiveness varied depending on the task type. Our results revealed that GroupfeedBot elicited even contribution among members, especially in the open-debating task (Figure 4). Among the various types of discussions, open debate on social issues is critical for a sound democracy [44]. An open debate has no single correct answer, and it is important to use a democratic process in which community members can all provide their arguments. In this aspect, it could be expected that a facilitating chatbot could at least minimally contribute to the development of participatory discussions.

Moreover, the participants who discussed with the facilitator chatbot perceived the discussion as efficient and the chatbot agent as useful only in the decision-making and open-debating tasks and not in the estimation and problem-solving tasks (Table 2). This suggests that GroupfeedBot's facilitating features could be more effective for tasks that require collaboration in which the members synthesize and mediate diverse opinions than for tasks involving finding the optimal solution from divergent ideas. Thus, chatbots should be properly applied to group discussions depending on the task's nature.

Future Work and Limitations

Here, we present limitations and future research directions. First, the sample size is small and does not sufficiently verify our findings. To generalize our preliminary results, further experiments should be conducted on a larger number of groups. Second, although we compared GroupfeedBot with the basic chatbot, we did not compare the facilitator chatbot with a human facilitator. Third, we used synchronous lab experiment situations rather than natural asynchronous situations. In the future work, we will apply the facilitator chatbot in a real-world setting without imposing any space–time constraints in the future work. Forth, when implementing the main features, we used one specific method (e.g., keyword presentation for summarization). Further research could be performed on various ways of implementing the chatbot's main features. Fifth, we could consider a chatbot that can function as a group chat manager in a natural setting, with the help of AI technology.

CONCLUSION

This study explored the feasibility of using a chatbot agent to support group chat discussions. We implemented GroupfeedBot, a chatbot agent which manages discussion time, encourages even participation, and organizes diverse opinions. Our results revealed that the facilitator chatbot can produce diverse opinion exchanges and encourage even participation, resulting in positive group dynamics.

ACKNOWLEDGEMENTS

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean National Police Agency and the Ministry of Science and ICT for police research and development (NRF-2018M3E2A1081492). We would like to thank Korea Creative Content Agency (KOCCA) for funding that partially supported this project. We also would like to thank Joseph Seering for extensive feedback on drafts. Finally, we would like to thank our participants for contributing to this study.

REFERENCES

[1] 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.

[2] Sandeep Avula, Gordon Chadwick, Jaime Arguello, and Robert Capra. 2018. SearchBots: User Engagement with ChatBots During Collaborative Search. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*. ACM, 52–61.

[3] Edmond Awad, Sohan Dsouza, Richard Kim, Jonathan Schulz, Joseph Henrich, Azim Shariff, Jean-François Bonnefon, and Iyad Rahwan. 2018. The moral machine experiment. *Nature* 563, 7729 (2018), 59.

[4] Rawia Awadallah, Maya Ramanath, and Gerhard Weikum. 2012. Harmony and dissonance: organizing the people's voices on political controversies. In *Proceedings of the fifth ACM international conference on Web search and data mining*. ACM, 523–532.

[5] Clint A Bowers, Curt C Braun, and Ben B Morgan Jr. 1997. Team workload: Its meaning and measurement. In *Team performance assessment and measurement*. Psychology Press, 97–120.

[6] 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*. ACM, 1223–1235.

[7] Karen Church and Rodrigo De Oliveira. 2013. What's up with whatsapp?: comparing mobile instant messaging behaviors with traditional SMS. In *Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services*. ACM, 352–361.

[8] Elizabeth F Churchill and Sara Bly. 1999. Virtual Environments at Work: ongoing use of MUDs in the Workplace. In *ACM SIGSOFT Software Engineering Notes*, Vol. 24. ACM, 99–108.

[9] Mercia Coetzee, Annette Wilkinson, and Daleen Krige. 2016. Mapping the social media landscape: a profile of tools, applications and key features. (2016).

[10] Justin Cranshaw, Emad Elwany, Todd Newman, Rafal Kocielnik, Bowen Yu, Sandeep Soni, Jaime Teevan, and Andrés Monroy-Hernández. 2017. Calendar. help: Designing a workflow-based scheduling agent with humans in the loop. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 2382–2393.

[11] Laura A Dabbish and Robert E Kraut. 2006. Email overload at work: an analysis of factors associated with email strain. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*. ACM, 431–440.

[12] Erika Darics. 2010. Relational work in synchronous text-based CMC of virtual teams. In *Handbook of research on discourse behavior and digital communication: Language structures and social interaction*. IGI Global, 830–851.

[13] Robert Dorfman. 1979. A formula for the Gini coefficient. *The review of economics and statistics* (1979), 146–149.

[14] Mark Dredze, Hanna M Wallach, Danny Puller, and Fernando Pereira. 2008. Generating summary keywords for emails using topics. In *Proceedings of the 13th international conference on Intelligent user interfaces*. ACM, 199–206.

[15] 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.

[16] Nicolas Fay, Simon Garrod, and Jean Carletta. 2000. Group discussion as interactive dialogue or as serial monologue: The influence of group size. *Psychological science* 11, 6 (2000), 481–486.

[17] Adrienne Porter Felt, Serge Egelman, and David Wagner. 2012. I've got 99 problems, but vibration ain't one: a survey of smartphone users' concerns. In *Proceedings of the second ACM workshop on Security and privacy in smartphones and mobile devices*. ACM, 33–44.

[18] James S Fishkin and Robert C Luskin. 1999. Bringing deliberation to the democratic dialogue. In *The poll with a human face*. Routledge, 13–48.

[19] Claudia I Flores-Saviaga, Brian C Keegan, and Saiph Savage. 2018. Mobilizing the trump train: Understanding collective action in a political trolling community. In *Twelfth International AAAI Conference on Web and Social Media*.

[20] Lawrence R Frey. 2004. The symbolic-interpretive perspective on group dynamics. *Small group research* 35, 3 (2004), 277–306.

[21] John Gastil and James P Dillard. 1999. Increasing political sophistication through public deliberation. *Political Communication* 16, 1 (1999), 3–23.

[22] Jeffrey Heer and Maneesh Agrawala. 2008. Design considerations for collaborative visual analytics. *Information visualization* 7, 1 (2008), 49–62.

[23] Jeffrey Heer, Fernanda B Viégas, and Martin Wattenberg. 2007. Voyagers and voyeurs: supporting asynchronous collaborative information visualization. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 1029–1038.

[24] James D Herbsleb, David L Atkins, David G Boyer, Mark Handel, and Thomas A Finholt. 2002. Introducing instant messaging and chat in the workplace. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 171–178.

[25] Ann Hill Duin and Ray Archee. 1996. Collaboration via e-mail and internet relay chat: Understanding time and technology. *Technical Communication* 43, 4 (1996), 402–412.

[26] Shirley S Ho and Douglas M McLeod. 2008. Social-psychological influences on opinion expression in face-to-face and computer-mediated communication. *Communication research* 35, 2 (2008), 190–207.

[27] Sungsoo Ray Hong, Minhyang Mia Suh, Nathalie Henry Riche, Jooyoung Lee, Juho Kim, and Mark Zachry. 2018. Collaborative dynamic queries: Supporting distributed small group decision-making. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 66.

[28] Tianran Hu, Anbang Xu, Zhe Liu, Quanzeng You, Yufan Guo, Vibha Sinha, Jiebo Luo, and Rama Akkiraju. 2018. Touch Your Heart: A Tone-aware Chatbot for Customer Care on Social Media. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 415.

[29] R Mark Isaac and James M Walker. 1988. Group size effects in public goods provision: The voluntary contributions mechanism. *The Quarterly Journal of Economics* 103, 1 (1988), 179–199.

[30] Daniel Katz and Robert L Kahn. 1978. *The social psychology of organizations*. Vol. 2. Wiley New York.

[31] Ralph Katz. 1982. The effects of group longevity on project communication and performance. *Administrative science quarterly* (1982), 81–104.

[32] Norbert L Kerr and Steven E Bruun. 1981. Ringelmann revisited: Alternative explanations for the social loafing effect. *Personality and social psychology bulletin* 7, 2 (1981), 224–231.

[33] Vishesh Khanna, Senthil N Sambandam, Arif Gul, and Varatharaj Mounasamy. 2015. "WhatsApp" ening in orthopedic care: a concise report from a 300-bedded tertiary care teaching center. *European Journal of Orthopaedic Surgery & Traumatology* 25, 5 (2015), 821–826.

[34] Sara Kiesler, Jane Siegel, and Timothy W McGuire. 1984. Social psychological aspects of computer-mediated communication. *American psychologist* 39, 10 (1984), 1123.

[35] Da-jung Kim and Youn-kyung Lim. 2015. Dwelling places in KakaoTalk: understanding the roles and meanings of chatrooms in mobile instant messengers. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 775–784.

[36] Soomin Kim, Joonhwan Lee, and Gahgene Gweon. 2019. Comparing Data from Chatbot and Web Surveys: Effects of Platform and Conversational Style on Survey Response Quality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 86.

[37] Tobias Kowatsch, Marcia Nißen, Chen-Hsuan I Shih, Dominik Rüegger, Dirk Volland, Andreas Filler, Florian Künzler, Filipe Barata, Severin Hung, Dirk Büchter, and others. 2017. Text-based healthcare chatbots supporting patient and health professional teams: Preliminary results of a randomized controlled trial on childhood obesity. *Persuasive Embodied Agents for Behavior Change (PEACH2017)* (2017).

[38] Mestan Küçük. 2010. Lurking in online asynchronous discussion. *Procedia-Social and Behavioral Sciences* 2, 2 (2010), 2260–2263.

[39] Bin Lin, Alexey Zagalsky, Margaret-Anne Storey, and Alexander Serebrenik. 2016. Why developers are slacking off: Understanding how software teams use slack. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion*. ACM, 333–336.

[40] Kiel Long, John Vines, Selina Sutton, Phillip Brooker, Tom Feltwell, Ben Kirman, Julie Barnett, and Shaun Lawson. 2017. Could you define that in bot terms?: Requesting, creating and using bots on reddit. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 3488–3500.

[41] Arnold M Lund. 2001. Measuring usability with the use questionnaire12. *Usability interface* 8, 2 (2001), 3–6.

[42] Moira McGregor, Nicola J Bidwell, Vidya Sarangapani, Jonathan Appavoo, and Jacki O'Neill. 2019. Talking about Chat at Work in the Global South: An Ethnographic Study of Chat Use in India and Kenya. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 233.

[43] Rada Mihalcea and Paul Tarau. 2004. Textrank: Bringing order into text. In *Proceedings of the 2004 conference on empirical methods in natural language processing*. 404–411.

[44] David Miller. 1992. Deliberative democracy and social choice. *Political studies* 40, 1_suppl (1992), 54–67.

[45] Seong-Jae Min. 2007. Online vs. face-to-face deliberation: Effects on civic engagement. *Journal of Computer-Mediated Communication* 12, 4 (2007), 1369–1387.

[46] Bonnie A Nardi, Steve Whittaker, and Erin Bradner. 2000. Interaction and outeraction: instant messaging in action. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work*. ACM, 79–88.

[47] Joaquin Navajas, Tamara Niella, Gerry Garbulsky, Bahador Bahrami, and Mariano Sigman. 2018. Aggregated knowledge from a small number of debates outperforms the wisdom of large crowds. *Nature Human Behaviour* 2, 2 (2018), 126.

[48] Les Nelson, Christoph Held, Peter Pirolli, Lichan Hong, Diane Schiano, and Ed H Chi. 2009. With a little help from my friends: examining the impact of social annotations in sensemaking tasks. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 1795–1798.

[49] Ani Nenkova and Amit Bagga. 2004. Facilitating email thread access by extractive summary generation. In *Proceedings of the 2003 International Conference on Recent Advances in Natural Language Processing (RANLP)*. 287–296.

[50] Midas Nouwens, Carla F Griggio, and Wendy E Mackay. 2017. WhatsApp is for family; Messenger is for friends: Communication Places in App Ecosystems. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 727–735.

[51] Annie Piolat, Thierry Olive, and Ronald T Kellogg. 2005. Cognitive effort during note taking. *Applied cognitive psychology* 19, 3 (2005), 291–312.

[52] Anabel Quan-Haase, Joseph Cothrel, and Barry Wellman. 2005. Instant messaging for collaboration: A case study of a high-tech firm. *Journal of Computer-Mediated Communication* 10, 4 (2005), JCMC10413.

[53] Joseph Seering, Juan Pablo Flores, Saiph Savage, and Jessica Hammer. 2018. The Social Roles of Bots: Evaluating Impact of Bots on Discussions in Online Communities. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 157.

[54] Joseph Seering, Michal Luria, Geoff Kaufman, and Jessica Hammer. 2019. Beyond Dyadic Interactions: Considering Chatbots as Community Members. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 450.

[55] In-geon Shin, Jin-min Seok, and Youn-kyung Lim. 2019. Ten-Minute Silence: A New Notification UX of Mobile Instant Messenger. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 442.

[56] Jane Siegel, Vitaly Dubrovsky, Sara Kiesler, and Timothy W McGuire. 1986. Group processes in computer-mediated communication. *Organizational behavior and human decision processes* 37, 2 (1986), 157–187.

[57] Ella Tallyn, Hector Fried, Rory Gianni, Amy Isard, and Chris Speed. 2018. The Ethnobot: Gathering Ethnographies in the Age of IoT. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 604.

[58] Carlos Toxtli, Andrés Monroy-Hernández, and Justin Cranshaw. 2018. Understanding chatbot-mediated task management. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 58.

[59] Jacob O Wobbrock, Leah Findlater, Darren Gergle, and James J Higgins. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 143–146.

[60] Anbang Xu, Zhe Liu, Yufan Guo, Vibha Sinha, and Rama Akkiraju. 2017. A new chatbot for customer service on social media. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 3506–3510.

[61] 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), 196.

[62] Sacha Zyto, David Karger, Mark Ackerman, and Sanjoy Mahajan. 2012. Successful classroom deployment of a social document annotation system. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1883–1892.