# GremoBot: Exploring Emotion Regulation in Group Chat

**Zhenhui Peng**
Hong Kong University of Science and Technology
zpengab@connect.ust.hk

**Taewook Kim**
**Xiaojuan Ma**
tw.kim@connect.ust.hk
mxj@cse.ust.hk
Hong Kong University of Science and Technology

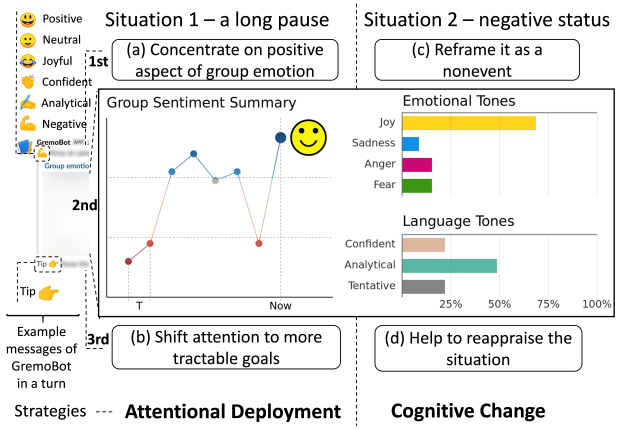


**Figure 1: Design of messages posted by GremoBot in the group chat. The 1st and 3rd messages are adapted from emotion regulation strategies [6] (i.e., attentional deployment in situation 1, cognitive change in situation 2). The 2nd message is visualized group emotion.**

## ABSTRACT

Maintaining a positive group emotion is important for team collaboration. It is, however, a challenging task for self-managing teams especially when they conduct intra-group collaboration via text-based communication tools. Recent advances in AI technologies open the opportunity of using chatbots for emotion regulation in group chat. However, little is known about how to design such a chatbot and how group members react to its presence. As an initial exploration, we design GremoBot based on text analysis technology and emotion regulation literature. We then conduct a study with nine three-person teams performing different types of collective tasks. In general, participants find GremoBot useful for reinforcing positive feelings and steering them away from negative words. We further discuss the lessons learned and considerations derived for designing a chatbot for group emotion management.

## INTRODUCTION

Maintaining a positive group emotion (e.g., every member is creative and active in a team project) is important for improving group relationships and performance [3, 4]. However, it is non-trivial especially when group collaboration is carried out online via the increasingly popular text-based group chat applications (e.g., Slack). For one thing, the formally appointed or emergent leaders qualified for group emotion regulation [13] are not always available in these group chats. For another, non-verbal cues are largely absent in group chat, making it harder to recognize the collective emotional states of group members from back-and-forth messages in rapid succession [16].

**Table 1: Script samples of GremoBot emotion regulation messages in Decision-Making (DM), Creativity (Cr), Debate (Db) tasks (task description in next page).**

| **Attentional Deployment Strategies** |
|---|
| (a) Concentration on positive aspect |
| (Positive)"The group has been making nice progress in the past few minutes." (Neutral)"Seems that the group discussion has been quite smooth so far." On dominant tone: (Joy)"I find the discussions valuable and enjoyable." (Confident)"Good to see the group's confidence is building up." (Analytical)"Glad that the group is taking an analytical approach to the problem." |
| (b) Shift attention to tractable goals |
| (DM) "How big is the candidate pool right now? Remember that the group needs to eventually narrow it down to five items." (Cr)"Feel free to share any idea, however small. The group can refine it together." (Db)"Try to express your opinions, as silence cannot help you win the debate." |
| **Cognitive Change Strategies** |
| (c) Reframe it as a nonevent |
| "I am sure that the group will work it out." |
| (d) Help to reappraise the situation |
| (DM)"You all have provided useful information that helps build the big picture. Compromise has to be made for the whole team's survival." (Cr) "It is a good start with everything the group has shared so far. Perhaps think outside the box and be adventurous." (Db)"The key to a convincing argument is to make other people feel comfortable enough to change their position." |

[1]GremoBot is open-source in https://github.com/PenguinZhou/GremoBot

Recent advances in artificial intelligence show the possibility of using a chatbot (or a bot) to automatically monitor group emotion and facilitate its regulation. Chatbots have been widely integrated into text-based chat applications, offering a variety of services such as information management [18] and task management [15]. However, few works explore its role as a moderator to manage emotion in a group chat [10]. In this work, we first follow theories in emotion regulation to design GremoBot (**Gr**oup **emo**tion **Bot**), a prototype that supports emotion regulation in Slack using commercial APIs. We then investigate how users react to GremoBot's assistance through an exploratory experiment with nine three-member groups on three types of collective tasks. Our contributions are: 1) present GremoBot[1], a proof-of-concept prototype for exploring the practical potential of using a chatbot to manage group emotion in a text-based teamwork environment; 2) provide an initial understanding of how group members perceive and work with GremoBot under different contexts, and 3) gain insights into the design of chatbot group emotion regulator.

## DESIGN OF GREMOBOT

GremoBot follows an external process of emotion regulation to monitor, evaluate and modify the group emotion [14], with the goal of encouraging group members to chat more positively.

**Monitoring via Text Analysis Technology.** For every text or emoji message $m_i$ in a group chat, GremoBot invokes the Microsoft Text Analytics API to get its sentiment score $s_i \in [0, 1]$ (1: most positive), and invokes the IBM Tone Analyzer API to detect its tones (i.e., emotional and language tones, Figure 1). As group emotion can be viewed as the sum of its member's affective compositions [2], we estimate the group sentiment over a time interval $T$ by averaging the sentiment scores of all the messages in that period, and measure tones distribution as the sum of all tones vectors divided by the total number of messages during $T$.

**Evaluation and Intervention.** We target on two situations: **1)** ***a long pause*** $T'$ $(< T)$ *in the group chat*, which may indicate a lack of progress in chats. GremoBot uses attentional deployment strategies to move members' attention towards group goals and the positive aspects of the current progress to encourage interaction (e.g., (a) and (b) in Figure 1). **2)** ***negative status****, i.e., negative group sentiment (score* $< 0.33$*) over the last time interval* $T$ *or accumulating* $n$ *negative messages in the past* $m$ $(> n)$ *posts.* GremoBot applies cognitive change strategies to help members interpret the situation in a positive way (e.g., (c) and (d) in Figure 1).

**Messages.** In each turn of intervention, GremoBot first interprets the situation positively, then visualizes group emotion to raise awareness, and finally gives a small tip for emotion regulation. Emojis are added to the chatbot messages to make them more vivid. Note that these text messages are pre-defined and customized by researchers in different tasks as exemplified in Table 1.

**Experiment Tasks:** 1) **D**ecision-**M**aking: from desert survival task [8]; each member ranks own 5 items first; group needs to pick and rank 5 out from the total 15 items; personal goal (50%): put own items in the final list; group goal (50%): as optimal as the expert's ranking;
2) **Cr**eativity: from [17]; come up at least 8 activities to support environmental sustainability; $5000 budget; judge group performance by number and feasibility of ideas;
3) **De**b**b**ate: promote a personal computer E-business; members hold different options: a) popular stars, b) computer game contest, c) entertainment TV shows; judge performance by each other.

**Experiment Procedure:** 1) for participants in each group, obtain consent; 2) introduce Slack, GremoBot, experiment; 3) in each task, 3 mins for reading materials and preparing; 4) 15 mins for discussion; 5) after each task, fill a questionnaire (adapted from [12]) about perceived usefulness of GremoBot in making them aware of group emotion and in giving tips to chat more positively, about annoyance (7-point Likert Scale), and about their main attention on the messages (multiple choices); 6) after all tasks, interview about user behaviors and design suggestions; and 7) debrief and compensation.

**More Details:** 1) 27 Participants (13 females, 14 males, labeled from P1-1 to P9-3) from graduate school in our university; age: 21 - 30 ($M$ = 23.81, $SD$ = 2.08); 2) Only use English or emoji (note: input as emoji name into the server) to chat in a typical Slack Windows app interface displayed on a Dell 2418H 24-Inch monitor; 3) Participants of a group sit in different desks and can not see each other during the chat; and 4) Participants with top-5 final performance can get extra bonuses.

## EXPERIMENT

To explore how people perceive and work with GremoBot, we conduct an exploratory study with nine three-member groups, each performing three types of group tasks (counter-balanced). As shown in the sidebar, the tasks cover scenarios in which each member has to consider both individual and group goals (**D**ecision-**M**aking), or mostly group goals (**Cr**eativity), or mostly individual goals (**De**b**b**ate). Through a pilot study with three additional groups, we experimentally set $T = 90s, n = 3, m = 10$ in all three tasks, and $T' = 31s$ in DM, $T' = 42s$ in Cr, $T' = 30s$ in Db task. Experiment procedure and more details are presented in the sidebar.

## RESULTS AND DISCUSSIONS

**Overall performance and perceptions.** In total, GremoBot actively intervened to regulate group emotion for 138 times due to the long pause ($M = 15.3, SD = 6.7$ per group), and 55 times due to the negative status ($M = 6.1, SD = 2.8$). It was triggered the most in the DM task (91 = 48 + 33), then in the Db task (84 = 71 + 13), and the least in the Cr task (28 = 19 + 9). We run a one-way repeated measures ANOVA with the task as the independent variable (using Bonferroni post-hoc test) on participants' perception of GremoBot (Table 2). Although the general ratings for GremoBot's usefulness are relatively low, there is a marginally significant difference on "I found it useful in making me aware of potential negative group emotion" ($F(2, 52) = 2.87, .05 < p < .1, \eta^2 = .10$), where it is significantly more useful in the Db task than in the DM task ($p < .05$). There is no significant difference across the three tasks regarding the perceived GremoBot's usefulness in suggesting participants to chat more positively ($F(2, 52) = 0.63, p = .54, \eta^2 = .024$). In terms of annoyance, a significant difference is found with respect to task type ($F(2, 52) = 5.81, p < .01, \eta^2 = .18$) - GremoBot is significantly more annoying in the DM task than in the Cr task ($p < .01$). These results suggest that GremoBot was perceived to be more useful in Cr- or Db-style group tasks than in DM tasks. *"In the first task (Db) its messages are important for reminding me to say something, but in the second task (DM) I am not even aware of its existence but just focus on the problem"* (P2-1, male, age: 22).

**Effects on user behaviors: 1) Reinforce the positives.** A smiley face, an increasing line in the visualization, and the dominantly joyful tone can have a positive reinforcement effect on the feeling and behavior of some participants. *"If I see the line jumps from the lower point to higher point, I am more happy and relaxed"* (P4-2, female, age: 24). *"One time I noticed the joyful group tone in the figure, I feel that I can use some joyful emoji to convey my happiness and make the chat fun"* (P1-3, male, age: 24).
**2) Alert members of negative situations.** GremoBot's intervention can warn some participants of the negative group emotion trend, triggering them to use more positive words or pushing them to chat more actively. *"I would check the line chart when it pops up, and if it goes down, I will speak (type) more quickly. It is a good impulse"* (P6-3, male, age: 24).

**Table 2: Perceived usefulness of GremoBot in making participants aware of negative emotion and chat more positively, as well as perceived annoyance in each task (1 - strongly disagree, 7 - strongly agree). Note: mean (SD),** $* : p < .05, ** : p < .01$.

| Task | Perceived Usefulness | | Perceived |
|---|---|---|---|
| | Aware emotion | Chat positively | Annoyance |
| DM | **3.3 (1.4)** * | 3.7 (1.5) | **4.0 (1.8)** ** |
| Cr | 3.7 (1.7) | 4.0 (1.7) | **3.0 (1.6)** ** |
| Db | **4.0 (1.4)** * | 3.9 (1.4) | 3.6 (1.4) |

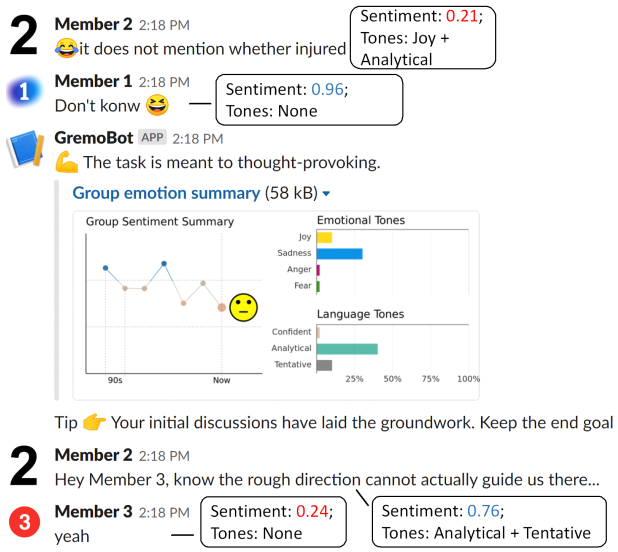


**Figure 2: An example of a chat screen shows that the participants' perception is mismatched with the GremoBot's evaluation. GremoBot detects negative sentiments (red scores), as illustrated by the labels in the text boxes.**

**Design Issues and Considerations: 1) Visualization of group emotion.** Most participants mainly paid attention to the group sentiment summary (count = 42) and group tones (count = 29) when GremoBot intervened. While the current emotion summary figure is easy to understand, it may break the flow in the group chat channel and make the bot less human-like. *"Sometimes I want to check previous messages in the first task (Cr), but it just pops out, and those messages move up"* (P3-1, female, age: 23). We suggest that this figure can be placed at a specific position on the chatting application interface to provide continuous feedback (as [9] did). **2) Emotion Regulation Strategy.** There are only 23 cases in which participants mainly pay attention to the textual emotion regulation messages, which are perceived useful in directing group dynamics but are not salient enough for users to notice most of the time. *"Its text messages are less attractive, compared to the large figure. But when I look back the whole group chat, I find that they could be extremely useful, like reminding us how many ideas we should come up with and asking us to extend existing ones"* (P5-2, male, age: 26). With the figure putting aside, we can further shorten the messages to reduce processing load by only using verbs and short phrases (e.g., "Good start! Now think outside the box!") to emphasize the main ideas. **3) Timing.** While the two situations for GremoBot's interventions are considered proper by participants, the exact timing should be adaptive to the group chat pace and task nature. *"If we speak fast, it should intervene less frequently. If we slow down later, it can pop up to remind us"* (P4-1, female, age: 24). In tasks like Decision-Making task, the bot should be more tolerant of negative words because *"group members need to analyze any negative outcome for better decision making"* (P5-2, male, age: 26). **4) Accuracy.** P2-3 mentioned that there is a mismatch between the group emotion they sensed and what was reported by GremoBot (Figure 2). *"I think it is not correct, especially in the second task (DM) and the third task (Cr). It reveals that our group emotion is going down, but we do not think so"* (P2-3, female, age: 21). In fact, the short sentences in group chat could be a big problem for GremoBot (even for human [5]), especially if we take culture into consideration [7]. We suggest that the chatbot for group emotion regulation should follow the guidelines for human-AI interaction [1], e.g., "make clear how well the system can do what it can".

## CONCLUSION AND FUTURE WORKS

In this paper, we designed GremoBot and conducted an exploratory study to evaluate its impacts on user perception and behavior. Results suggest that a chatbot emotion regulator can enhance positive feelings and alert people of negative situations. We further discuss design issues raised in user feedback and propose design considerations concerning visual feedback, textual strategy, timing and accuracy of emotion regulation chatbots. We identify two future research directions: 1) improving the message design of GremoBot and conducting a controlled experiment (with vs. without) to systematically evaluate its feasibility and efficacy; and 2) investigating the appropriate manner of GremoBot in managing emotions, e.g., proactive intervention or reactive involvement [11].

## REFERENCES


[1] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 3, 13 pages. https://doi.org/10.1145/3290605.3300233

[2] Sigal Barsade and Donald Gibson. 1998. Group emotion: A view from top and bottom. *Research on Managing Groups and Teams* 1 (01 1998), 81–102.

[3] Sigal G. Barsade. 2002. The Ripple Effect: Emotional Contagion and its Influence on Group Behavior. *Administrative Science Quarterly* 47, 4 (2002), 644–675. https://doi.org/10.2307/3094912

[4] Sigal G. Barsade, Andrew J. Ward, Jean D. F. Turner, and Jeffrey A. Sonnenfeld. 2000. To Your Heart's Content: A Model of Affective Diversity in Top Management Teams. *Administrative Science Quarterly* 45, 4 (2000), 802–836. https://doi.org/10.2307/2667020 arXiv:https://doi.org/10.2307/2667020

[5] Alastair J. Gill, Darren Gergle, Robert M. French, and Jon Oberlander. 2008. Emotion Rating from Short Blog Texts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*. ACM, New York, NY, USA, 1121–1124. https://doi.org/10.1145/1357054.1357229

[6] James J. Gross. 1998. The Emerging Field of Emotion Regulation: An Integrative Review. *Review of General Psychology* 2, 3 (1998), 271–299. https://doi.org/10.1037/1089-2680.2.3.271

[7] Ari MJ Hautasaari, Naomi Yamashita, and Ge Gao. 2014. "Maybe It Was a Joke": Emotion Detection in Text-only Communication by Non-native English Speakers. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 3715–3724. https://doi.org/10.1145/2556288.2557215

[8] J. Clayton Lafferty and Patrick M. Eady. 1974. The desert survival problem. *Plymouth, Michigan: Experimental Learning Methods* (1974).

[9] 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 (CHI '09)*. ACM, New York, NY, USA, 537–546. https://doi.org/10.1145/1518701.1518784

[10] Moodbit. 2019. https://mymoodbit.com/ Accessed: 2019-04-03.

[11] Zhenhui Peng, Yunhwan Kwon, Jiaan Lu, Ziming Wu, and Xiaojuan Ma. 2019. Design and Evaluation of Service Robot's Proactivity in Decision-Making Support Process. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 98, 13 pages. https://doi.org/10.1145/3290605.3300328

[12] Zhenhui Peng, Jeehoon Yoo, Meng Xia, Sunghun Kim, and Xiaojuan Ma. 2018. Exploring How Software Developers Work with Mention Bot in GitHub. In *Proceedings of the Sixth International Symposium of Chinese CHI (ChineseCHI '18)*. ACM, New York, NY, USA, 152–155. https://doi.org/10.1145/3202667.3202694

[13] Anthony T Pescosolido. 2002. Emergent leaders as managers of group emotion. *The Leadership Quarterly* 13, 5 (2002), 583–599.

[14] Ross A. Thompson. 1994. Emotion Regulation: A Theme in Search of Definition. *Monographs of the Society for Research in Child Development* 59, 2/3 (1994), 25–52. http://www.jstor.org/stable/1166137

[15] 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 (CHI '18)*. ACM, New York, NY, USA, Article 58, 6 pages. https://doi.org/10.1145/3173574.3173632

[16] JOSEPH B. WALTHER. 1992. Interpersonal Effects in Computer-Mediated Interaction: A Relational Perspective. *Communication Research* 19, 1 (1992), 52–90. https://doi.org/10.1177/009365092019001003

arXiv:https://doi.org/10.1177/009365092019001003

[17] Teng Ye and Lionel P. Robert, Jr. 2017. Does Collectivism Inhibit Individual Creativity?: The Effects of Collectivism and Perceived Diversity on Individual Creativity and Satisfaction in Virtual Ideation Teams. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. ACM, New York, NY, USA, 2344–2358. https://doi.org/10.1145/2998181.2998261

[18] Amy X. Zhang and Justin Cranshaw. 2018. Making Sense of Group Chat Through Collaborative Tagging and Summarization. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 196 (Nov. 2018), 27 pages. https://doi.org/10.1145/3274465