Robot Gaze Can Mediate Participation Imbalance in Groups with Different Skill Levels

Sarah Gillet*
Ronald Cumbal*
sgillet@kth.se
ronaldcg@kth.se
KTH Royal Institute of Technology
Stockholm, Sweden

André Pereira
atap@kth.se
KTH Royal Institute of Technology
Stockholm, Sweden

José Lopes
jd.lopes@hw.ac.uk
Heriot-Watt University
Edinburgh, United Kingdom

Olov Engwall
engwall@kth.se
KTH Royal Institute of Technology
Stockholm, Sweden

Iolanda Leite
iolanda@kth.se
KTH Royal Institute of Technology
Stockholm, Sweden

ABSTRACT

Many small group activities, like working teams or study groups, have a high dependency on the skill of each group member. Differences in skill level among participants can affect not only the performance of a team, but also influence the social interaction of its members. In these circumstances, an active member could balance individual participation without exerting direct pressure on specific members by using indirect means of communication, such as gaze behaviors. Similarly, in this study, we evaluate whether a social robot can balance the level of participation in a language skill-dependent game, played by a native speaker and a second language learner. In a between-subjects study (N = 72), we compared an adaptive robot gaze behavior, that was targeted to increase the level of contribution of the least active player, with a non-adaptive gaze behavior. Our results imply that, while overall levels of speech participation were influenced predominantly by personal traits of the participants, the robot's adaptive gaze behavior could shape the interaction among participants which lead to more even participation during the game.

KEYWORDS

Multiparty interaction, gaze, group dynamics, language learning

ACM Reference Format:
Sarah Gillet, Ronald Cumbal, André Pereira, José Lopes, Olov Engwall, and Iolanda Leite. 2021. Robot Gaze Can Mediate Participation Imbalance in Groups with Different Skill Levels. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (HRI '21), March 8–11, 2021, Boulder, CO, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3434073.3444670

*Both authors contributed equally to this research.

HRI '21, March 8–11, 2021, Boulder, CO, USA

ACM ISBN 978-1-4503-8289-2/21/03…$15.00
https://doi.org/10.1145/3434073.3444670

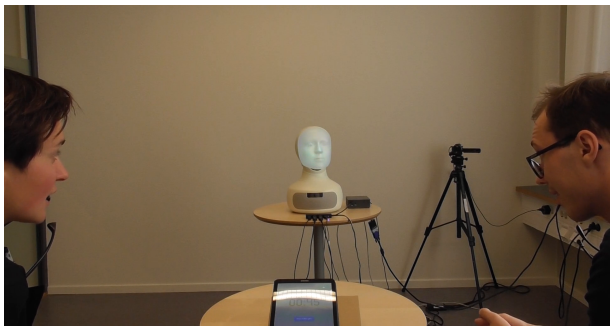


Figure 1: Overview of the interaction between the Furhat robot and two participants with different skill levels in a skill-based task.

1 INTRODUCTION

Many of our everyday interactions take place in small groups. Although diversity in terms of knowledge and skills possessed by each group member may cause groups to outperform individuals, when the skill gap between members becomes too large this may prevent the group from exploiting this potential [39]. Especially in educational settings, with learners at different skill levels (e.g., domain knowledge, linguistic proficiency, etc.), creating a prosperous learning environment in small group activities remains a difficult task, even in traditional classroom situations [6].

Social robots have previously been successful in educational settings [1, 30] and group interactions [4, 7, 17, 23]. Furthermore, social robots have been able to influence group dynamics such as cohesion [33] and inclusion [34] and support groups in situations of conflict [15, 29] and debate [12, 25]. Group dynamics is a complex factor that is constantly evolving [36] which calls for adaptive robot behaviors.

As 60-65% of human communication is non-verbal [2], the exploration of non-verbal robot behaviors for influencing groups might be promising. Non-verbal robot behavior has, previously, been used to mediate participation in groups. In a recent work, a specially designed microphone-shaped robot balanced participation in a conversation by actively encouraging participants through non-verbal

behaviors [37]. We explore how adaptive non-verbal robot behaviors, specifically gazing, can balance participation in groups with diverging skill levels.

To investigate how natural, adaptive, non-verbal robot behaviors could support the development of group dynamics, we chose language as a skill and paired a native speaker and a second language learner of the same language to play a language-focused game with a robot. The different robot behaviors were evaluated in a between-subject design. Seventy-two participants in thirty-six pairs engaged in a game in which players had to describe target words while another player guessed which word was described. In our experiment, the fully autonomous robot took the role of the guesser and the pair of participants cooperated to provide relevant verbal information that allowed the robot to guess the word (see Figure 1). As gaze behavior has been identified as a promising method to influence groups [21, 31], we designed the robot's gaze behaviors to autonomously adapt moment-by-moment using real-time evaluation of participation levels with the goal of structuring and balancing participation in the activity.

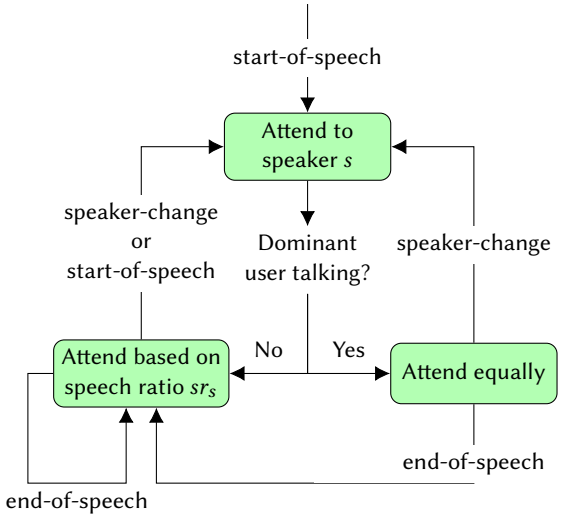


Figure 2: Diagram describing the flow of the robot gaze behavior.

2 RELATED WORK

To investigate how a robot could maximize a team's productivity, Claure et al. [3] utilized Multi-Armed Bandits managed by a robot system in a collaborative Tetris game so that the robot could assign resources based on the team members' skill levels. As this results in the less skilled member receiving fewer resources, they further examined how this could be achieved while still maintaining perceived fairness in the distribution of resources among team members. Our work explores the same phenomenon of different group skills, but we focus on promoting long-term improvement in skill for the individual and the group rather than the team's immediate productivity. In social educational settings, the application of robots as facilitators in a collaborative group task has been found promising in terms of time management, objectivity, and efficiency [26].

An increasing number of human-robot interaction (HRI) studies have focused on how robots can influence group dynamics. These works can be divided into *directly* or *indirectly* influencing methods that use *verbal* or *non-verbal* approaches. Considering direct verbal methods, a robot was found to be able to influence group dynamics in conflict situations by openly addressing interpersonal violations [15]. Similarly, verbal robot interventions were used to resolve object possessing conflicts among children constructively [29]. With the goal of enhancing human-human collaboration among children, a robot that expressed interpersonal cohesiveness compared to task cohesiveness utterances was found to lead to higher perception of team performance [32]. Further, Shamekhi and Bickmore [28] showed how a robot could act as a facilitator and improve human-human meetings. As one aspect of meeting facilitation, they explored a direct verbal method to influence participation equality by asking passive participants about their opinion before accepting the group decision. Other authors have studied the effects of indirect verbal behaviors (i.e., not addressing the targeted group dynamics directly). For example, by making vulnerable expressions, e.g., "Sorry guys, I made the mistake this round", a ripple effect was found in the team of three humans and the robot that lead to more trust-related behaviors within the group [35] and improved the dynamics of the conversation [38]. Further, robots have shown to improve inclusion of out-group team members, both for adult teams [34] and groups of children [9].

Methods that indirectly influence group dynamics use non-verbal behaviors such as gaze and backchannels. In pioneering research, Mutlu and colleagues [20, 21] showed how gaze behavior can shape conversation roles by manipulating the frequencies with which the robot gazed at participants. Gaze also has been shown to increase story recall [19] and the persuasive power of a robot [10]. Further, Skantze [31] analysed how a pair of participants interacting with a Furhat robot reacted to the robot giving the turn to one randomly chosen participant. These results indicated that participants successfully took the turn from a humanoid robot following its gaze cues. As one of the most related examples to our work, the microphone-shaped MicBot robot [37] was designed with the goal of balancing conversation participation through direct non-verbal behaviors. By encouraging the least active participant by turning towards them, MicBot could achieve higher and more even engagement in a conversation among a group of three. Our work builds on the results from this study and on the large body of research of how nonverbal behavior can influence human group interactions. We investigate how a human-like robot embodiment with natural adaptive gaze behaviors could encourage the less active participant in a game and thereby balance participation, in the particular case of groups with different skill levels.

3 ADAPTIVE GAZE BEHAVIOR FOR BALANCING GROUP PARTICIPATION

In this work, we developed an adaptive robot gaze behavior that reacts to the current estimate of participation in the task, which is used as an indication for the dynamics of the group. Based on findings by Mutlu et al. [20, 21], the robot shifts its attention between

the users with the aim of shaping participant roles and encourage participation by the least active participant in cases of silence. Figure 2 gives an overview of how the robot adapts its gaze behavior based on the current group behavior. If the participant with the larger amount of speaking time (the dominant participant) is talking, the robot attends to both participants equally. If the participant with the smaller speech amount (non-dominant participant) is talking, the robot instead shifts gaze based on the relative amounts of speaking time. The latter gaze-shift behavior is also executed in cases of silence. We define the relative speech amount, or more precisely the speech ratio sr_p, for participant p as

$$sr_p = \frac{\text{speech}_p}{\text{speech}_o}, \tag{1}$$

where speech_p is the accumulated time in seconds of speech of participant p and speech_o that of the other participant o. To achieve an encouraging and speech balancing behavior, we use sr_p to define the proportion of the robot's gazing towards either participant. The idea is that the gaze is distributed evenly when the dominant participant is talking, but that the non-dominant participant gets a share of gaze time that is $\frac{1}{sr_p}$ larger than the other participant when talking. Algorithm 1 shows how the gazing behavior is calculated. The gaze-shifting decision making algorithm is called whenever a change in talker occurs and is then repeated whenever a gaze change is intended (after *timeNextAction* seconds) and is always performed with the current real-time estimations of sr_p for each participant.

While developing the robot behavior, we further considered that humans might not fully align their head angle with their gaze, particularly for small gaze angles [8]. As only eye gaze, though, might not be noticed and the coordination of movement and gaze was shown to be important [40], the robot executes small head rotations towards the participant who is the gaze target. These head rotations are coordinated with and support the eye gaze. The robot's head pose is further determined by the current speech ratio and, therefore, the current group dynamic. The smaller sr_p is for the currently talking participant (or in case of silence), the more the head pose will point towards this participant. If the dominant participant is talking the head pose will point to the middle between participants.

Algorithm 1: Gaze-shifting decision making

Data: p_a = participant at whom the robot is currently gazing
t_p = seconds the current gaze lasted
minimumGaze = minimum time for a gaze,
maximumGaze = maximum time for a gaze
Result: Participant to gaze at, time for gaze, head pose
p_t = participant who is currently talking;
p_{nt} = participant not currently talking;
sr_{p_t} = compute speech ratio following equation 1;
if $p_a == p_t$ **then**
  // gaze at participant currently not talking
  ratio = min(sr_{p_t}, 1);
  timeNextAction = $\text{ratio}^2 * t_p$;
  headPose = calculate head pose from speech ratio;
  return p_{nt}, timeNextAction, headPose;
else
  // gaze at current talker
  // for a speech ratio of 1.0, keep given times for gaze to sample from, otherwise shorten gaze if ratio > 1.0, lengthen gaze if ratio < 1.0
  timeInfluence = 1.0 - sr_{p_t};
  timeInfluence = max(timeInfluence, (minimumGaze - maximumGaze) / 3.0);
  timeNextAction = sampleBetween(minimumGaze + timeInfluence, maximumGaze + timeInfluence) ;
  headPose = calculate head pose from speech ratio;
  return p_t, timeNextAction, headPose
end

4 STUDY DESIGN

To evaluate the effects of the adaptive gaze behavior on human group dynamics, we designed a between-subject study with two conditions. In the experimental condition, the robot behaves as described in the previous section; in the control condition the robot instead gazes mostly only at the speaker (more details in Section 4.2). As a task, we designed a variant of the game *With Other Words* (original in Swedish *Med Andra Ord*) in which two human participants can simultaneously give hints to a robot that attempts to guess which word was hinted at by using an autonomous guess generator (c.f. Section 4.3.1). In each group, the pair of participants was formed by a native Swedish speaker and a second language learner of Swedish with low-intermediate to advanced language proficiency.

4.1 Hypotheses

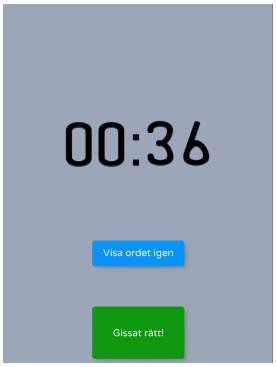
As discussed in Section 2, Mutlu et al. [21] suggested that a robot is able to shape conversational roles through gaze and Skantze [31] showed that participation could be induced by non-verbal behaviors coupled with gaze. Based on these findings, we formulate our first two hypotheses:

- **H1** *Participants will participate more evenly in describing the words when playing with a robot showing gaze behaviors that attempt to balance group participation.*
- **H2** *Participants will speak more and take more turns when interacting with a robot that encourages balanced participation.*

As the robot is displaying a gaze behavior that encourages less active participants and particularly language learners to practice their skill with a native speakers, we further hypothesize that:

- **H3** *The robot will be perceived as more socially present when balancing the interaction through gaze.*
- **H4** *Considering language learners, the willingness to practice the language will increase throughout the interaction when interacting with the experimental condition.*

(a) The game displaying the target word on the tablet.



(b) The game screen after the word is hidden.

Figure 3: The game *With Other Words* as shown on the tablet to the participants. The word in the example is *kyrktorn* - church tower. First, participants are shown both the word and the picture below the timer (3a). After a predefined amount of time, the word and picture are hidden but can be shown again upon request (3b). Tablet buttons: *Visa ordet igen* - Show word again; *Gissat rätt!* - Guessed correctly!

4.2 Conditions

To test our hypotheses, we compared the robot gaze behavior as described in Section 3 in the *experimental* condition with an alternative believable gaze behavior in the control condition. As the robot in the *control* condition was not considering real-time evaluation of group dynamics, the robot was always gazing at the person currently talking to attend to the flow of the conversation. This behavior was realized by using the same voice activation detection as in the experimental condition. As the gazing behavior in the experimental condition is dynamic in gaze targets, we employed a similar dynamic in the control condition. The robot changed its gaze target with a similar frequency as the robot in the experimental condition, but it was performing gaze aversion (with targets left or right above the participant, dependent on seating) instead of gazing at the other participant. The robot performed gaze aversion ~25% of the time, meaning if the robot had looked 5s at the speaker, it would perform gaze aversion for 1.25s and then look back to the speaker. The gaze target for aversion was kept constant to avoid the impression of a robot randomly looking around in the room. Because gazing to the other participant could recreate the behavior present in the *experimental* condition, we decided to not gaze at the other, non-talking participant, in the *control* condition. Guessing the words and other non-verbal expressions (i.e., backchannelling) were performed in the same way in both conditions.

4.3 With Other Words

Our goal with using a variant of the *With Other Words* game was to create an engaging task in which the different proficiency skills among participants would become apparent while ensuring that spoken language practice would still be possible for participants with low proficiency. In each interaction, participants collaboratively tried to produce utterances that describe a target word shown on a tablet (see Figure 3) to the robot. As in the original game, the target guess word was forbidden and could not be pronounced.

Participants were instructed that they could describe each target word together and did not need to take separate turns for each round. Each round with one target guess word had a maximum play time of 60 seconds or was terminated when the participants indicated that the robot guessed the word correctly by clicking a button in the tablet. We provided an image representing the meaning of the word to avoid the situation in which language learners could become passive because they did not know the meaning of the word. Both the target word and accompanying descriptive image were hidden after 8 seconds to avoid participants focusing their attention on the tablet instead of the robot, as the participants could then risk missing the gaze and head pose patterns of the robot. Similar to the original game, participants played words with three increasing difficulty levels. The higher the difficulty, the more the skill – language – is needed. The reason for using different difficulty levels was to study the influence of the robot behavior as the skill (in our case, language proficiency) becomes more critical for accomplishing the task of describing the guess word.

Participants had to play a minimum of 10 target words and 6 minutes on the *easy* level, 8 words and 6 minutes on the *medium* level, and 4 words and 3 minutes on the *hard* level. Participants were not aware of the difficulty level increase. The target words available in the game were randomly taken from the Swedish pocket version of the original game while ensuring their feasibility for the autonomous guessing process. During the game, the target words were randomly sampled from the current difficulty level. To avoid possible confounds with a specific word order, there was no given order or word list that was played by each participant pair.

4.3.1 Autonomous guess generation. The process of generating guesses in a human-like manner during the game is complex given the various ways in which the participants could explain each target word (e.g., descriptions, completing a sentence, synonyms, situational similarities). To reduce this complexity, we limited the guessing vocabulary knowledge of the system, i.e., for each target word, we defined a list of useful alternative words – the guessing vocabulary. These lists of words were defined during extensive pilot studies in which the robot's guessing was performed by a wizard (i.e., controlled by a human operator). To decide which word from this list would be the most appropriate, the automatic speech recognition system[1] continuously recorded the participants' speech (with a separate microphone each) and processed individual transcriptions. The received transcriptions were then filtered for keywords. The guessing process then aimed at computing the similarity between keywords in the description and each word of the guessing vocabulary. Therefore, the process of prompting a guess started by computing Word2Vec[18][2] embeddings of all keywords given by the participants and encode them in a single average context vector. Afterwards, we compared the distance between the system's guess vocabulary and the context vector with cosine similarity. The vocabulary word with the least distance to the context vector was chosen as a guess. To decide when the correct word should be uttered by the robot, we used the data of the extensive pilot studies

[1] Google Cloud Speech API
[2] Swedish CoNLL17 corpus from http://vectors.nlpl.eu/repository

and analyzed when, given the length of descriptions in seconds, the wizard would have the robot utter the correct guess word. In the process of guessing, the robot reacted to pauses in the speech activation. If the speech recognition system detected a predefined amount of keywords in the transcribed speech (in our case 20%) and the sample based on the distribution of length of descriptions indicates that the guess should be given, the robot uttered the word. For example, if participants described the word for ~7 - ~20 seconds, there would be a 20% chance of the robot guessing the word in the next pause.

We noticed during the pilot studies that giving the correct guess while gazing at one participant might be perceived as a direct reaction to this participant's last utterance and influence that participant's confidence level (and future participation in the game). To avoid this potential confound with our experimental manipulation, the robot looked at the middle of the two participants when saying the correct guess word. We did not observe that incorrect guesses uttered by the robot would influence confidence. Therefore, we only made this exception to the robot gaze behavior when the robot was uttering the correct guess.

The autonomous guessing process was monitored by a wizard. Whenever the participants were talking and therefore did not hear or understand correctly what the robot said, the wizard could repeat the word. Caused by a misinterpretation of the picture, participants in some rare cases accidentally described a different word. In these cases, the wizard had the robot utter the word participants were expecting. The wizard intervened for the two latter cases combined for ~2% of all uttered words.

4.3.2 The system set-up. We used a Furhat robot[3] with the iRobot face texture and the Swedish female voice *Elin* from Acapela group[4]. The game interface implemented in Unity was displayed on a tablet. Each participant was wearing two close-talk microphones, one for voice activation detection and one for automatic speech recognition.

4.4 Measurements

We collected a variety of measures during the game as well as additional measures in a pre- and post-experiment questionnaire.

4.4.1 In-game measures. As part of the study design, the words participants were asked to describe had three different increasing difficulty levels. To reflect the influence of the robot's behavior as the language proficiency (imbalanced skill between participants) became more critical, we collected each of the following characteristics for each difficulty level in the game.

Active participation - During the game, we measured the **amount of active participation** by analyzing the duration of voice activity detection (VAD) for that participant. For each difficulty level, the active amount is the total duration of VAD in relation to the total duration of that difficulty level.

Number of turns taken - We measured how often the participants were taking turns. With two participants, the number of turns one can take is dependent on the other participant taking their turns, so we defined this measure on a group level. For each difficulty level and group, the number of turns taken was normalized to the duration of that difficulty level in minutes.

Unevenness of participation - Following [37], we obtain the difference between each individuals' speech time and the average speech time for the group. These differences are then accumulated to measure the group's deviation from the mean and express how uneven the group's speech was distributed. This measure results in values close to 0 if the group distributes speech evenly (which we consider a good approximation for balanced participation in the game). Higher numbers reflect an uneven distribution. We used the following formula:

$$\text{uneven}_{\text{speech}} = \sum |s_i - \bar{s}| \tag{2}$$

with s_i representing the speech of participant i relative to the total speech in the pair and $\bar{s}$ the mean of the pair to compute the **unevenness of participation**. Note that as the s_i of the group sum to 1 due to the normalization on total speech, the mean always results in a value of 0.5. This measure therefore reflects how far the group deviates from the expected even speech distribution among participants (each participant speaks 50% of the time).

4.4.2 Questionnaire measures. The following measures were collected in pre- and post experiment questionnaires.

Personality traits - We selected the dimensions of extroversion and agreeableness from the Big Five Inventory [13, 14] as we expected that these two personality traits would have the highest influence on participation behavior and therefore our task.

Familiarity - Inspired by Strohkorb Sebo et al. [35], we measured self-reported familiarity between participants. To measure this familiarity, a five-point Likert scale ranging from "I have never met the other participant" to "I am close friends with the other participant" is used. Further, we asked about connections on social media and if they had each other's phone numbers. The latter two added one point each to the scale of familiarity when answered positive.

Language proficiency - The self-reported proficiency of Swedish was collected on the scale of language levels of the Common European Framework of Reference for Languages [22], with the addition of "Native speaker".

Social presence - In order to investigate how the manipulation of the gaze behavior was perceived by participants, we employed two scales of the Networked Minds Questionnaire [11] targeted at the robot on a Seven-point Likert scale: *Co-presence* is defined as the degree to which the observer believes s/he is not alone and *attentional allocation* describes the amount of attention the user allocates to and receives from an interactant. The questionnaire was shown to be applicable to human-robot interaction scenarios [16].

Willingness to communicate - We used a questionnaire validated by Ryan [27] to assess the general motivation of a person to communicate in a second language.The questionnaire comprises 8 questions about how likely one would be to communicate in a given situation in the given language, i.e., Swedish. The questions are answered on a six-point Likert scale. This measure is used to evaluate the willingness to practice the second language and will be used to evaluate hypothesis **H4**.

[3] https://furhatrobotics.com/
[4] https://www.acapela-group.com/

Figure 4: Overview over the experimental set-up.

4.5 Experimental procedure

After giving written consent, participants completed the pre-experiment questionnaire comprising items about the demographics of participants, the two dimensions of personality traits and their proficiency in Swedish individually. The game instructions were included in the consent form (in English) and were verbally repeated (in Swedish) by the robot at the beginning of the interaction. Participants were instructed that they could both describe the target word. Participants, one native Swedish speaker and one Swedish language learner, were randomly assigned to one of the two study conditions. After filling the pre-questionnaire, they were guided to the meeting room with the Furhat robot and asked to take a seat in front of it. The chairs were placed with a ~60 degree angle and 2 meters distance between each other (see Figure 4). The experimenter explained the microphone placement and left the room before Furhat started introducing itself. Then the robot asked participants to state their name and where they came from so that each of them would be aware of the possible difference in language skill. After this introduction, the robot repeated the game rules and participants interacted with the Furhat robot and the game for 15-20 minutes. After completing the game with the robot, participants were guided to individual rooms again to complete the post-questionnaire, which comprised the familiarity between participants, social presence and willingness to communicate. Participants were thanked for their participation with a voucher (value ~9 USD).

4.6 Participants

In total, 72 participants were recruited to interact with the Furhat robot in 36 pairs. Participants were recruited from the surrounding town and the university campus through flyers, posters, word of mouth and online platforms. Recruitment material contained the information that participants would interact with another participant and the Furhat robot. Participants' age ranged between 18 and 67 years ($M = 31.43$, $SD = 10.77$) and 35 identified as female, 36 as male and 1 did not say. Participants were paired randomly according to their role (1 native speaker, 1 language learner). Besides 1 participant who reported a medium familiarity with the other participant, participants were unfamiliar with each other. 39 of the participants had never interacted with a robot before and 19 reported to interact with robots regularly.

Level	Easy		Medium		Hard	
	M	SD	M	SD	M	SD
Experimental	0.30	0.15	0.25	0.11	0.25	0.16
Control	0.41	0.24	0.35	0.25	0.4	0.30

Table 1: Means and standard deviations of the unevenness of participation for the conditions at each difficulty level.

5 RESULTS

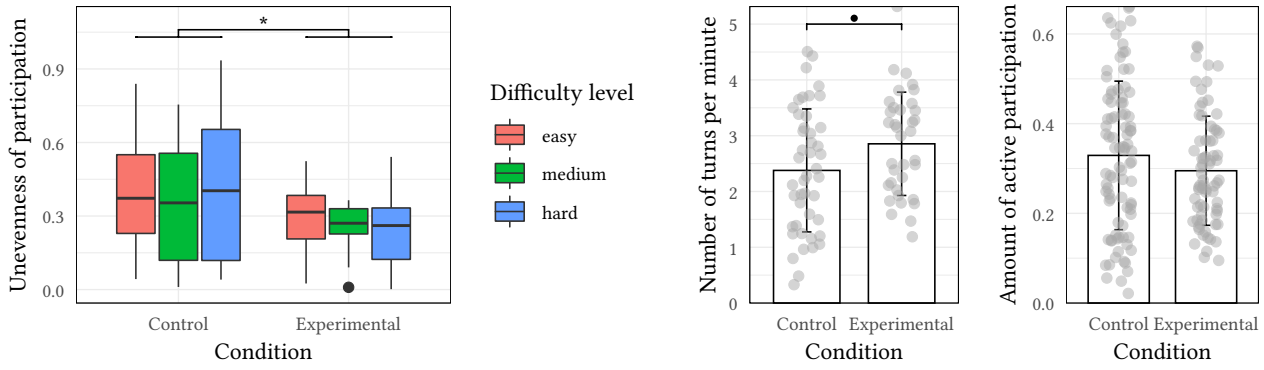
Of the 36 groups, 9 groups were excluded from analyses either because of substantial hardware and software problems (8 groups) or misinterpretation of instructions by the participants (1 group). Of the remaining 27 groups (54 participants) that were included in the analysis, 12 groups interacted in the experimental condition (12 female, 11 male, 1 rather not say) and 15 in the control condition (12 female, 18 male) with an average age of 31.79 years ($SD = 11.50$) and 32.03 years ($SD = 10.73$), respectively. The proficiency level of language learners was distributed between Low-intermediate (1 learner, 3 learners), Intermediate (5 learners, 4 learners), High-intermediate (4 learners, 6 learners) and Advanced (2 learners, 2 learners) for experimental and control condition, respectively. Participants played the game and interacted with the Furhat robot for an average of 15.8 minutes ($SD = 0.37$).

5.1 Unevenness of participation

To evaluate the impact of our manipulation, we performed a two-way ANCOVA and examined the effects of *condition* and *level of difficulty* on the *unevenness of participation*, after controlling for the *proficiency of the language learner*. We chose *proficiency of language learners* as a covariate because we expected that language learners according to their proficiency would struggle more in describing the words. *Level of difficulty* was chosen as a predictor as we expected that language learners would have more difficulty describing the words with increasing difficulty levels. The analysis yielded a main effect of condition for the unevenness of participation measure, $F(1, 72) = 5.717$, $p = 0.019$ whilst controlling for *proficiency of language learners*, meaning that the robot's behavior could influence the participation behavior and lead to less uneven participation in the experimental condition ($M = 0.27, SD = 0.14$) compared to the control condition ($M = 0.39, SD = 0.26$). The effect of the difficulty level was not significant and the covariate language proficiency was not significantly related to the unevenness of participation. Figure 5a gives an overview over the *unevenness of participation* between conditions and difficulty levels. The means and standard deviations for the respective interactions can be found in Table 1.

5.2 Number of turns taken

To analyze the effects of our study manipulation on the *number of turns taken*, we performed a two-way ANCOVA in the same way and examined the effects of *condition* and *level of difficulty* on *number of turns taken*, after controlling for the *proficiency of the language learner*. The covariate, *language proficiency*, was significantly related to *number of turns taken*, $F(3, 72) = 2.819$, $p = 0.045$ indicating that the proficiency of the language learners influenced the number turns taken in the group. Detailed inspection of the data revealed



(a) Unevenness of participation per difficulty level and condition. Lower numbers signal more even participation. The graph shows the progression of the interaction over time in light of the different difficulty levels. *p<0.05

(b) Average number of turns and amount of active participation per condition. Error bars represent the standard deviation.•p<0.1

Figure 5: Results from the analysis of unevenness of participation, number of turns and amount of active participation.

that the higher the proficiency of the language learner was, the more turns were taken in the group. The main effect of condition showed a trend to significance, $F(1, 72) = 3.500, p = 0.065$, indicating that, as illustrated in Figure 5b, groups showed a trend to taking more turns in the experimental condition ($M = 2.85, SD = 0.93$) compared to the control condition ($M = 2.38, SD = 1.10$). The effect of difficulty level was not significant.

5.3 Amount of active participation

A two-way ANCOVA was performed to examine the effects of *condition* and *level of difficulty* on the *amount of active participation*, after controlling for *willingness to communicate*, *language proficiency*, and the personality traits of *agreeableness* and *extroversion*. As the personality traits of extroversion and agreeableness impact how talkative and cooperative a person is [5], we chose these personality traits as covariates. Further, language proficiency affects how capable a person is at communicating and the willingness to communicate comprises how likely a person is to use the language in different situations, which led to treating the measures of *language proficiency* and *willingness to communicate* as covariates. As we were performing an analysis at the individual level, we added "Native speaker" as the highest proficiency level, resulting in five different proficiency levels.

The covariate, extroversion, was significantly related to the speech amount, $F(1, 151) = 8.192, p = 0.005$, indicating after further inspection of the data that higher levels of extroversion led to higher amounts of participation. There was also a significant relation of proficiency to the speech amount, $F(4, 151) = 13.333, p < 0.001$ indicating that the proficiency level of the participant influenced the amount of active participation. The predictors condition and difficulty level were not significant. This means that the robot behavior did not have a significant influence on the amount of general participation ($M = 0.3, SD = 0.12$ for experimental condition, $M = 0.33, SD = 0.17$ for control condition, see Figure 5b). Each participant did not talk for an average 68.5% of the game, which included time when the other participant was talking, when the robot was trying to guess the word and silence. Often, participants were silent as they thought about how to describe the target word or waited for the robot to say more guess words.

5.4 Questionnaire measures

Additionally, we analyzed the influence of the robot gaze behavior on the perception of social presence. Linear models for co-presence and attentional engagement with *condition* as predictor did not yield significant results. The scale of co-presence showed a trend towards significance, $F(1, 52) = 3.043, p = 0.087, R^2 = 0.055, R^2_{adjusted} = 0.037$, and was generally rated high on the seven-point Likert scale with $M = 6.472, SD = 0.677$ and $M = 6.133, SD = 0.734$ for the experimental and control conditions, respectively. The analysis of the willingness to communicate comparing pre- and post-experiment measures did not yield significant results.

6 DISCUSSION

The goal of this study was to evaluate whether the gaze behavior of a robot could influence participation levels of human group members (prompting them to participate equally) with different skills (i.e., language proficiency: one native speaker and one language learner).

6.1 Shaping of group interactions

The results support **H1** as participation was significantly more balanced in the groups assigned to the experimental condition. This indicates that robots can use natural gaze behaviors to help group members participate evenly in an activity despite possible differences in skill level. Figure 5a further shows how the participation imbalance is shaped through time based on the condition. Where in the control condition the results show a high variation between groups, the experimental condition leads to less variation and a slight improvement over time despite the increase in difficulty.

Even though our results did not support **H2**, predicting an increased amount of participation and turn taking, the significant relation between proficiency and amount of participation found in

the analysis undertaken for **H2** can serve as manipulation check. Our manipulation (i.e., skill imbalance) was successful as different skill levels lead to significantly different amounts of active participation in the skill-based task.

Combining the support for **H1**, indicating more balanced participation, with the significant relation between proficiency and amount of participation found when evaluating **H2**, the results indicate that even though the robot could not influence an individual's amount of participation, it could shape how the group was interacting by balancing the expected unevenness of participation induced by the skill imbalance in the group. With this result, we extend prior literature that explored balancing engagement among group members through a non-antromorphic robot [37] to balancing participation through a conversational robot and gaze behaviors. Different from prior work, we purposefully manipulated the dynamic in the group by inducing skill imbalance. Further, our algorithmic approach to gaze behavior is grounded in prior work by Mutlu et al. [20, 21] that explored how gaze patterns can be used to shape static roles in a conversation. In our work, the robot autonomously observes participation levels in the group and decides how to shape the roles accordingly. Thereby, we shed light on how a robot could utilize shaping roles dynamically depending on the current situation in the group to achieve a more balanced group interaction.

6.2 Perception of the robot

Additionally, the balancing behavior shown by the robot seemed not to be perceived by participants as far as influencing their perception of co-presence or attentional engagement as the results for both measures were non-significant and hypothesis **H3** could not be supported. This, on one hand, indicates that the chosen control condition offers a fair comparison to the experimental condition and, on the other hand, is in line with prior literature showing that the gaze behaviors executed on a robot influence participants behavior in a more subtle way [24, 40]. The willingness to communicate among language learners did not change throughout the interaction resulting in **H4** not being supported. As the short amount of time might be a reason for the absent change, future work could explore if longer interactions with encouraging and participation balancing gaze behaviors could increase the willingness to practice the skill.

6.3 Implications based on the task

As a last part of the post-experiment questionnaire, we offered participants to note any comments or observations about the robot or the interaction. As our task could also be considered as a state-of-the art learning problem in natural language processing, participants solely focused on the way the robot guessed in the open comments and described that it was a really good – sometimes maybe too good – guesser. Only one participant in an early pilot mentioned the gazing behavior as being encouraging. This indicates that state-of-the-art problems, even when being simplified for the purpose of an HRI study (as many participants indicated that the robot might be cheating) offer an interesting task for studying subtle group influencing behaviors.

Our task was chosen because it specifically requires language skills and therefore naturally reveals differences in skill if one of the players is not a native speaker. Therefore, the proposed task implements our intended manipulation by providing the robot with the opportunity to perceive the imbalances in the participants' behaviors and to aim for balancing them. At the same time, the game requires a second set of skills regarding the creativity to describe words which is independent of language proficiency. Therefore, we expect that our results can be transferred to situations beyond educational tasks in which team success actually requires diverse skill sets but the social structure or individual differences can hinder participation of some members. The exploration of those situations is left to future work.

6.4 Limitations

The robot's behavior in both conditions was dependent on the automatic voice activation captured by the headset microphones worn by each participant. We also based most of our analysis on this automatic collection. As these microphones might catch background noise, these measures are expected to have inaccuracies. Nonetheless, this noise and, therefore, the inaccuracies are expected to be present in both conditions in the same way. Further, our study was conducted in a group with two participants with different skill levels and one robot. Future work has to explore if the results could be extended to groups with more participants or more robots. Additionally, it remains open for future work if the explored robot behavior could be extended to groups with other kinds of imbalances in group dynamics (e.g., based on personality traits).

7 CONCLUSION

Prior work has shown that robots can use gaze to influence groups. We extend these prior works by showing the applicability of human-like robot gaze behaviors and their influence on group constellations with skill imbalance and therefore specific need for mediation. We developed an autonomous robot gaze behavior that could balance group members' participation by adapting to the real-time perceived participation. While personality and language proficiency had a strong influence on the amount of participation, the robot could shape how participants interacted with each other by balancing participation levels. Our results offer possible applications to robot facilitators for meetings or educational activities in which different skill levels are common, to ensure even participation through the robot's gaze behavior.

8 ACKNOWLEDGEMENTS

We thank Sanne van Waveren, Ilaria Torre, and Catherine Weldon for their support in preparing this paper and the unit of language and communication at KTH for their support in recruiting participants. This work was partially funded by grants from the Swedish Research Council (no. 2017-05189, no. 2016-03698), the Swedish Foundation for Strategic Research (FFL18-0199) and the Jacobs Foundation (no. 2017 1261 06).

REFERENCES

[1] Tony Belpaeme, James Kennedy, Aditi Ramachandran, Brian Scassellati, and Fumihide Tanaka. 2018. Social robots for education: A review. *Science robotics* 3, 21 (2018).

[2] Judee K. Burgoon, David B. Buller, Jerold L. Hale, and Mark A. de Turck. 1984. Relational Messages Associated With Nonverbal Behaviors. *Human Communication Research* 10, 3 (1984), 351–378. https://doi.org/10.1111/j.1468-2958.1984.tb00023.x

[3] Houston Claure, Yifang Chen, Jignesh Modi, Malte Jung, and Stefanos Nikolaidis. 2020. Multi-Armed Bandits with Fairness Constraints for Distributing Resources to Human Teammates. (2020), 299–308. https://doi.org/10.1145/3319502.3374806

[4] Filipa Correia, Samuel Mascarenhas, Rui Prada, Francisco S. Melo, and Ana Paiva. 2018. Group-based Emotions in Teams of Humans and Robots. *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction - HRI '18* February (2018), 261–269. https://doi.org/10.1145/3171221.3171252

[5] Paul T Costa and Robert R McCrae. 1992. *Revised NEO personality inventory (NEO-PI-R) and Neo five-factor inventory (NEO-FFI).* Psychological Assessment Resources.

[6] Jennifer Evans and Ingrid Lunt. 2002. Inclusive education: are there limits? *European Journal of Special Needs Education* 17, 1 (2002), 1–14. https://doi.org/10.1080/08856250110098980 arXiv:https://doi.org/10.1080/08856250110098980

[7] Marlena R. Fraune, Selma Šabanović, and Takayuki Kanda. 2019. Human group presence, group characteristics, and group norms affect human-robot interaction in naturalistic settings. *Frontiers Robotics AI* 6, JUN (2019), 1–16. https://doi.org/10.3389/frobt.2019.00048

[8] JH Fuller. 1992. Comparison of head movement strategies among mammals. *The headneck sensory motor system. Oxford University Press, New York* (1992), 101–112.

[9] Sarah Gillet, Wouter van den Bos, and Iolanda Leite. 2020. A social robot mediator to foster collaboration and inclusion among children. In *Proceedings of Robotics: Science and Systems*. Corvalis, Oregon, USA. https://doi.org/10.15607/RSS.2020.XVI.103

[10] Jaap Ham, René Bokhorst, Raymond Cuijpers, David Van Der Pol, and John John Cabibihan. 2011. Making robots persuasive: The influence of combining persuasive strategies (gazing and gestures) by a storytelling robot on its persuasive power. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 7072 LNAI (2011), 71–83. https://doi.org/10.1007/978-3-642-25504-5{_}8

[11] Chad Harms and Frank Biocca. 2004. Internal consistency and reliability of the networked minds social presence measure. *Seventh Annual International Workshop: Presence 2004. Valencia: Universidad Politecnica de Valencia.* (2004).

[12] Guy Hoffman, Oren Zuckerman, Gilad Hirschberger, Michal Luria, and Tal Shani Sherman. 2015. Design and Evaluation of a Peripheral Robotic Conversation Companion. *ACM/IEEE International Conference on Human-Robot Interaction* 2015-March (2015), 3–10. https://doi.org/10.1145/2696454.2696495

[13] Oliver P John, Eileen M Donahue, and Robert L Kentle. 1991. The big five inventory—versions 4a and 54.

[14] Oliver P John, Laura P Naumann, and Christopher J Soto. 2008. Paradigm shift to the integrative big five trait taxonomy. *Handbook of personality: Theory and research* 3, 2 (2008), 114–158.

[15] 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*. Association for Computing Machinery, Portland, Oregon, USA, 229–236. https://doi.org/10.1145/2696454.2696460

[16] Iolanda Leite, Carlos Martinho, Andre Pereira, and Ana Paiva. 2009. As time goes by: Long-term evaluation of social presence in robotic companions. In *RO-MAN 2009-The 18th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 669–674.

[17] Iolanda Leite, Marissa McCoy, Daniel Ullman, Nicole Salomons, and Brian Scassellati. 2015. Comparing Models of Disengagement in Individual and Group Interactions. *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction - HRI '15* March (2015), 99–105. https://doi.org/10.1145/2696454.2696466

[18] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*. 3111–3119.

[19] Bilge Mutlu, Jodi Forlizzi, and Jessica Hodgins. 2006. A storytelling robot: Modeling and evaluation of human-like gaze behavior. *Proceedings of the 2006 6th IEEE-RAS International Conference on Humanoid Robots, HUMANOIDS* (2006), 518–523. https://doi.org/10.1109/ICHR.2006.321322

[20] Bilge Mutlu, Takayuki Kanda, Jodi Forlizzi, Jessica Hodgins, and Hiroshi Ishiguro. 2012. Conversational gaze mechanisms for humanlike robots. *ACM Transactions on Interactive Intelligent Systems* 1, 2 (2012), 1–33. https://doi.org/10.1145/2070719.2070725

[21] Bilge Mutlu, Toshiyuki Shiwa, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. 2009. Footing in human-robot conversations: how robots might shape participant roles using gaze cues. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*. ACM, 61–68.

[22] Council of Europe. Council for Cultural Co-operation. Education Committee. Modern Languages Division. 2001. *Common European Framework of Reference for Languages: learning, teaching, assessment.* Cambridge University Press.

[23] Raquel Oliveira, Patrícia Arriaga, Patrícia Alves-Oliveira, Filipa Correia, Sofia Petisca, and Ana Paiva. 2018. Friends or Foes? Socioemotional Support and Gaze Behaviors in Mixed Groups of Humans and Robots. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction - HRI '18*, Vol. 9. ACM Press, New York, New York, USA, 279–288. https://doi.org/10.1145/3171221.3171272

[24] André Pereira, Catharine Oertel, Leonor Fermoselle, Joseph Mendelson, and Joakim Gustafson. 2020. Effects of Different Interaction Contexts When Evaluating Gaze Models in HRI. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (Cambridge, United Kingdom) *(HRI '20)*. Association for Computing Machinery, New York, NY, USA, 131–139. https://doi.org/10.1145/3319502.3374810

[25] Danielle Rifinski, Hadas Erel, Adi Feiner, Guy Hoffman, and Oren Zuckerman. 2020. Human-human-robot interaction: robotic object's responsive gestures improve interpersonal evaluation in human interaction. *Human–Computer Interaction* 00, 00 (2 2020), 1–27. https://doi.org/10.1080/07370024.2020.1719839

[26] Rinat B Rosenberg-Kima, Yaacov Koren, and Goren Gordon. 2020. Robot-Supported Collaborative Learning (RSCL): Social Robots as Teaching Assistants for Higher Education Small Group Facilitation. *Front. Robotics and AI* 2020 (2020).

[27] Stephen Ryan. 2009. Self and identity in L2 motivation in Japan: The ideal L2 self and Japanese learners of English. *Motivation, language identity and the L2 self* 120 (2009), 143.

[28] Ameneh Shamekhi and Timothy W. Bickmore. 2019. A multimodal robot-driven meeting facilitation system for group decision-making sessions. *ICMI 2019 - Proceedings of the 2019 International Conference on Multimodal Interaction* (2019), 279–290. https://doi.org/10.1145/3340555.3353756

[29] Solace Shen, Petr Slovak, and Malte F. Jung. 2018. " Stop . I See a Conflict Happening ." A Robot Mediator for Young Children ' s Interpersonal Conflict Resolution. In *Human Robot Interaction*. 69–77.

[30] Michihiro Shimada, Takayuki Kanda, and Satoshi Koizumi. 2012. How Can a Social Robot Facilitate Children's Collaboration?. In *International Conference on Social Robotics*. Springer, Berlin, Heidelberg, 98–107. https://doi.org/10.1007/978-3-642-34103-8{_}10

[31] Gabriel Skantze. 2017. Predicting and Regulating Participation Equality in Human-robot Conversations: Effects of Age and Gender. *ACM/IEEE International Conference on Human-Robot Interaction* Part F1271 (2017), 196–204. https://doi.org/10.1145/2909824.3020210

[32] Sarah Strohkorb, Ethan Fukuto, Natalie Warren, Charles Taylor, Bobby Berry, and Brian Scassellati. 2016. Improving human-human collaboration between children with a social robot. *25th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2016* (2016), 551–556. https://doi.org/10.1109/ROMAN.2016.7745172

[33] Sarah Strohkorb, Iolanda Leite, Natalie Warren, and Brian Scassellati. 2015. Classification of Children's Social Dominance in Group Interactions with Robots. *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction - ICMI '15* (2015), 227–234. https://doi.org/10.1145/2818346.2820735

[34] Sarah Strohkorb Sebo, Ling Liang Dong, Nicholas Chang, and Brian Scassellati. 2020. Strategies for the Inclusion of Human Members within Human-Robot Teams. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. 309–317.

[35] Sarah Strohkorb Sebo, Margaret Traeger, Malte F. Jung, and Brian Scassellati. 2018. The Ripple Effects of Vulnerability: The Effects of a Robot's Vulnerable Behavior on Trust in Human-Robot Teams. *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction - HRI '18* February (2018), 178–186. https://doi.org/10.1145/3171221.3171275

[36] Michael Sweet and Larry K Michaelsen. 2007. How group dynamics research can inform the theory and practice of postsecondary small group learning. *Educational Psychology Review* 19, 1 (2007), 31–47.

[37] Hamish Tennent, Solace Shen, and Malte Jung. 2019. Micbot: A Peripheral Robotic Object to Shape Conversational Dynamics and Team Performance. *ACM/IEEE International Conference on Human-Robot Interaction* 2019-March (2019), 133–142. https://doi.org/10.1109/HRI.2019.8673013

[38] Margaret L Traeger, Sarah Strohkorb Sebo, Malte Jung, Brian Scassellati, and Nicholas A Christakis. 2020. Vulnerable robots positively shape human conversational dynamics in a human–robot team. *Proceedings of the National Academy of Sciences* (3 2020), 201910402. https://doi.org/10.1073/pnas.1910402117

[39] Daan Van Knippenberg, Carsten KW De Dreu, and Astrid C Homan. 2004. Work group diversity and group performance: An integrative model and research agenda. *Journal of applied psychology* 89, 6 (2004), 1008.

[40] Marynel Vázquez, Elizabeth J. Carter, Braden McDorman, Jodi Forlizzi, Aaron Steinfeld, and Scott E. Hudson. 2017. Towards Robot Autonomy in Group Conversations. *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction - HRI '17* (2017), 42–52. https://doi.org/10.1145/2909824.3020207