Teamwork and Social Robots: Examining Forced Participation Within Group Endeavors

Evelyn Roberts

Advised by Sarah Sebo, 5th Year PhD Candidate in Computer Science, and Brian Scassellati, Professor of Computer Science, Cognitive Science, and Mechanical Engineering

Submitted to the faculty of Cognitive Science in partial fulfillment of the requirements for the degree of Bachelor of Science

Yale University
April 22, 2019

Abstract

In addition to being useful in many different settings, robots can be used to aid in specific social settings or learn about human development. Prior research has shown that interaction with a robot, through both verbal and nonverbal interactions, can aid team settings and group dynamics (Matsuyama et al. 2015, Correia et al. 2018, Tennent, Shen & Jung, 2019). The focus of this study is to determine what effect a social robot can have on breaking down faultlines, or divisions that form, between individuals in a group within a controlled setting. We specifically hope to examine how forced participation, through designation of a "team captain" position to either an in-group or out-group member of a larger group, can potentially influence how the group is able to function as a cohesive unit while completing a two-part timed survival task. Results of the present experiment suggest that forced participation by one participant, particularly an outsider participant, significantly affects (decreases) group talking time. Means of participant self-reports of equal treatment and warmth of the robot also significantly varied with whether an outsider or insider was the group team captain. These results potentially suggest that all types of group participation are not equally effective as interventions for dealing with faultlines, and that artificially trying to engage all members of a group through interaction with a social robot can lead to noticeable (and arguably negative) behavioral performance in participants, in addition to changed perceptions of group equality and robot character traits.

Table of Contents

1 Introduction

1.1 Teamwork

Group behavior, which includes teamwork, is integral to many aspects of human life, and has also been a topic of focus in recent years within academic and workforce domains. Studies conducted by Wooley et al. show results supporting the existence of a "collective intelligence"(c) of groups that can affect group performance, suggesting that group success is a function of both individual qualities and total group dynamic: this collective intelligence is theorized as being a combination of individual group member's general intelligence (g) and social sensitivity, or ability to work with others (Wooley et al., 2010).

Much research has been conducted on the success of teams within formal spheres: in their analysis, Sundstrom, De Meuse, and Futrell examine how teams can be used effectively within organizations (such as companies) and other domains (such as sports teams and committees). A work team is defined as an "interdependent [collection] of individuals who share responsibility for specific outcomes for their organizations", and the success of these groups is affected by many key factors, including organizational culture, mission clarity, autonomy, and technology and task design (Sundstrom, Meuse & Futrell, 1990). The authors claim that the foundation of successful organizations are "self-regulating work teams", and that these teams oftentimes achieve peak performance through member cohesion, open-mindedness, and desire to continue working together.

1.2 Faultlines

Other literature suggests that faultlines, or the divisions that form between people, are an important factor to consider when discussing effective teamwork (Meyer et al. 2014, Lau & Murnighan 1998, 2005). Meyer et al. (2014) claim that discussing and acknowledging natural

faultlines is necessary, now more than ever, due to the increasing number of diverse teams developing in the modern workforce. Faultlines have also been shown to have a negative impact on team goals and outcomes whether they are noticeable to members of the team or not (Thatcher & Patel, 2012), suggesting that successful group dynamics could rely upon proper treatment of existing faultlines, among other considerations, whether these groups are explicitly aware of the divisions that form or not.

Lau and Murnighan define the phenomena of faultlines within their "Demographic Diversity and Faultlines: The Compositional Dynamics of Organizational Groups" while also diving into potential ways to address and/or mitigate the effects that faultlines can have on groups. A more recent collaboration between the two discusses faultlines, as well, although their 1998 article provides a theoretical discussion and their 2005 article an overview of a current experiment involving faultlines. Within their 1998 publication, the authors discuss how initial faultlines within a group usually fall along demographic lines, as these are oftentimes the most salient characteristics of a person upon first meeting. However, "more subtle [and varied] characteristics" become common sources for faultline alignment as time within the group goes on. Faultlines are also multidimensional and can be built upon many different characteristics: faultlines gain strength as more similar attributes cause participants to align themselves in the same way (Lau & Murnighan, 1998). An example of a single faultline would be gender where members of a group would be divided into smaller male and female subgroups. If all the women in the group were also white and all the men black, this would also create a strengthened single faultline informed by both gender and race. The size of subgroups created by faultlines is also very important to how groups function, particularly if different subgroups have significantly unequal sizes. When a majority subgroup is compared to a minority one, a power inequality

oftentimes forms, which can have large effects on the ways that groups perceive one another and choose to act. Members of minority subgroups are more likely to experience pushback or suppression when voicing opinions that do not align with the majority, and a larger group's functioning may seem smooth and without conflict to most group members even if there have been instances of suppression of the smaller subgroup (Lau & Murnighan, 1998).

1.3 Social Robots and Teamwork

In recent years, the use of robots in non-mechanical settings has increased greatly, most noticeably in the realm of healthcare and mental health (Riek, 2015), in-home companionship (Culturally competent robots), and childhood development (Crompton, Gregory & Burke, 2018). Social robots, or robot(s) that "interacts and communicates with humans or other autonomous physical agents by following social behaviors and rules attached to its role" (Social robot) can be used to aid in social development and help researchers examine human behavior, particularly group behavior. In their 2015 study, Matsuyama et al. explore how robots might aid in group engagement in four-participant groups, particularly through regulation of imbalanced engagement density (i.e. targeting participants with lower instances of participation). Their results show promising evidence of successful manipulation of group engagement that is positively received by group members (Matsuyama et al., 2015). In their study, Correia et al. also show how the use of group-based emotions by social robots in a multi-participant setting can aid in overall feelings of group trust and connectedness. Their study compared group dynamics within a group that had either a robot member who expressed individual or group-based emotions, with results showing that participants rated the robot who expressed group-based emotions as being more likeable, and additionally eliciting higher levels of group identification and trust (Correia et al., 2018) than the robot who expressed individual emotions.

Other studies have shown that successful group interaction with robots need not be verbal or acknowledged as being related to team dynamics. In their "Micbot: A Peripheral Robotic Object to Shape Conversational Dynamics and Team Performance", Tennent, Shen & Jung showcase the success that a non-humanoid robot—a peripheral robotic object—can have on group dynamics (Tennent, Shen & Jung, 2019). Through engaging movements, their robot was implicated in successfully helping increase group engagement and demonstrated problem-solving performance. This study shows the wide variety of success that different human-robot interaction interventions can have, and suggests that robots need not be explicitly human-like in order to positively affect group performance or mood: there are likely numerous other unexplored ways that team dynamics can be aided by human-robot interaction, a promising direction for the field as a whole.

1.4 Present Study

This present study was designed in relationship to another very similar control study also involving faultline manipulation within a group, a social robot, and a survival task. In the control study, backchanneling (an active listening skill where "verbal and nonverbal signals [are used]...to display [one's] attentiveness to speakers' utterances" (Dixon & Foster 1998)) was utilized by social robot Jibo in order to test how differential treatment might affect group dynamics and individual feelings about the task. Specifically, these concerns were addressed by measuring how talking time, influence over group decisions, and participant feelings were affected by this type of treatment from the robot. The present study was designed using the same survival task, general procedure, and robot, but altered the human-robot interaction element in order to examine a different aspect of human teamwork and teambuilding. This experiment examines how people work together and feel they are working with a group when one member of

the group is able to communicate unequally with a social robot in a room for necessary information: in one condition, a member of an artificially-created ingroup is asked to communicate unequally with a social robot, and in the other a member of an artificially-created outgroup is asked to do the same.

In this paper, we aim to explore one method of intervention in artificial faultline formation: forced question-asking (querying) interaction of one in-group or out-group member with a social robot in a group survival task setting. We see this forced participation as a way to expand upon prior literature on group dynamics and faultlines, and explore if types of interaction between ingroup and outgroup members within a larger group can be enhanced by use of social robots. As discussed in Lau & Murnighan's article, faultlines can cause rifts between groups, especially when subgroups are of different sizes, leading to power inequalities and altered group dynamics when members of the smaller (minority) group attempt to state their opinions or go against the majority will. Through our experimental design, we hope to explore generally how group dynamics are affected by the presence of a social robot, but also investigate another question: will minority subgroup members stating their opinion have the same detrimental effect on overall group dynamics when forced to do so by another entity (such as the social robot and experimenter) and more closely enforced by the presence and help of a social robot physically present in the room?

1.5 Hypotheses

We hypothesized that forced robot interaction among participants would be beneficial to overall group dynamics and predicted that this would manifest in a few different ways. Generally, we predicted that the outgroup member will feel more included in the group when they are chosen to interact with the robot, whereas members of the ingroup will feel less of a

difference between being forced to participate and not doing so. We hypothesized that behavior, perceptions of the robot, and perceptions of the group would be affected by forced participation with a social robot, namely that participants chosen as team captain from the outgroup would give more positive reports on these dimensions than when an ingroup member was chosen. We also hypothesized that these variables would be more positive for ingroup members when chosen as team captain than when not, but that this difference would not be as large as for the outsider chosen to talk to the social robot. (i.e. The difference in feelings of inclusion, behavior, etc. between an outgroup member being chosen as team captain or not will be larger than the difference between an ingroup member being chosen as team captain or not.)

2 Methods

2.1 Participants

We analyzed data from 45 participants age 18 and over who completed one of two conditions run for this experiment, designated TC-insider or TC-outsider. Participants were recruited from Yale University and the greater New Haven area. Sign up flyers were posted in coffee shops, university and residential college bulletin boards, and online spaces such as club rosters, Facebook groups, and email panlists for recruitment purposes.

2.2 Survival Task

The experiment was split into two parts, the first part 15 minutes and the second part 30 minutes long. Before beginning the experiment, participants were informed that they would have to collaborate with the other study participants in order to complete a two-part timed activity based off a popular survival task. In the first part of the experiment, participants completed the same task in both the individual and group rooms, where they were informed by Jibo (the social robot) that they were stranded in an unknown foreign place (with one other person in the group

room and alone in the individual room), surrounded by a few familiar household items. Participants were given a list of these 25 items, which included a coffee pot, floss, a cd, a teddy bear, and whiskey, among others. Within the 15 minutes, participants are asked to consider the items on the list with the help of Jibo (and their partner if within the group room) and rank order the items for importance to their survival in the unknown location from 1 to 25, 1 being the most important and 25 being the least. In this part of the task, Jibo was able to answer questions about the items but not the location in which they are stranded. If a participant states "Hey Jibo, tell me about the shoelaces", Jibo will respond: "Shoelaces: The shoelaces are each 3 and a half feet long and are neon yellow in color"; or asks about the garbage bag, "Garbage bag: 1 strong drawstring large trash bag that can hold up to 30 gallons of garbage. Black in color".

All three participants were brought together in the group room for the second part of the experiment and were able to use their materials (i.e. item lists and scratch paper) from the first part of the experiment to aid their analysis in the second part. In this section, Jibo reveals that its GPS has determined the location in which they are stranded and can now answer questions about the location and items using certain keywords such as "animals", "weather", and "temperature", and that only one participant will be able to ask Jibo questions related to items or environment. Participants are informed that they can only bring 8 items from their initial list of 25 now, and are asked to use the clues about the items and environment to come up with a revised list of the 8 most important items that they may use in order to survive in the location. When asked about something like the temperature in this part, Jibo can respond "Generally, the average temperature hovers around 13 degrees Celsius or 55 degrees Fahrenheit...". In this part of the experiment, Jibo can also still give the same answers about the item descriptions as in the first part.

2.2.1 Experimental Manipulation: TC-Insider versus TC-Outsider

The two conditions for this experiment were identical except for the identity of the person able to communicate with Jibo directly. For the TC-Insider condition, abbreviated as TC-in, the person able to communicate with Jibo directly (ask it questions about the location or items and have Jibo respond in a meaningful way) was a member of the insider group, or the group of 2 in the group room originally. They were always participant 1. In the TC-Outsider condition, abbreviated as TC-out, the person able to communicate with Jibo directly was a member of the outsider group, or the group of 1 in the individual room originally. They were always participant 3. All participants were informed that "only one person would be able to ask Jibo questions in this part of the task, and for you all this is [Participant Name]". The two conditions, in summary:

1. **Team Captain-Insider (TC-in)**: The participant able to ask Jibo questions directly is from the insider group, or the group of 2 in the group room in part 1 of the experiment.
2. **Team Captain-Outsider (TC-out)**: The participant able to ask Jibo questions directly is from the outsider group, or the group of 1 in the individual room in part 1 of the experiment.

2.3 Procedure

Instructions for each task were given to participants beforehand, and experimenters remained outside experiment rooms for any technical or logistical questions that might come up during the task itself. Participants were assigned to number 1, 2, or 3 randomly before the beginning of the experiment, and these numbers determined group membership, seating location, and ability to ask Jibo questions directly by condition (i.e. forced group participation). Designation of "team captain" was always done by the experimenter after Jibo had given instructions for the second part of the experiment. In the TC-insider condition, this person was participant 1, or the person sitting farthest to the right originally in the group of 2, and in the TC-

outsider condition this was participant 3, or the person sitting farthest to the left originally in the group of 1.

Pre-experiment surveys were administered to all three participants at the same time after they had signed a waiver (with information about the task and their involvement) and were given their participant ID numbers (1 - 3 with their group letter attached). Instructions were given by the experimenter, then groups were split up into randomly selected groups of 1 and 2 people. The group of 1 was set up in a back room first while the group of 2 stayed and asked each other questions from a sheet to get to know each other better. Examples of such questions were "Which emoji do you use the most and why?" and "If you didn't sleep, what would you do with your extra time?". Once the 1 person group was set up, the experimenter then set up the 2 person group in a different room (the group room) with the exact same task as within 1 person room. The experimenter and Jibo issued instructions, mics were tested, and then participants had 15 minutes for the first part of the task. After the 15 minutes passed for the group of 1, they were led out of the room to wait outside while the group of 2 finished part 1. When part 1 ended for both groups, the individual was led to the group room and handed a mic. All participants were then informed of the nature of the second part of the task by the experimenter and Jibo, mics were tested again for the individual asking Jibo questions, and then participants were left alone to complete their task. Before beginning, the group was notified about the identity of the group member able to directly ask Jibo questions. When finished, participants were led to another room to take their post-experiment surveys, with 1 chair apart from the other 2. Location of seating was recorded. Participants were paid $10 or $15 ($15 for the last 4 groups run) when they completed the post-experiment survey or upon completion of whatever tasks that participants felt comfortable completing.

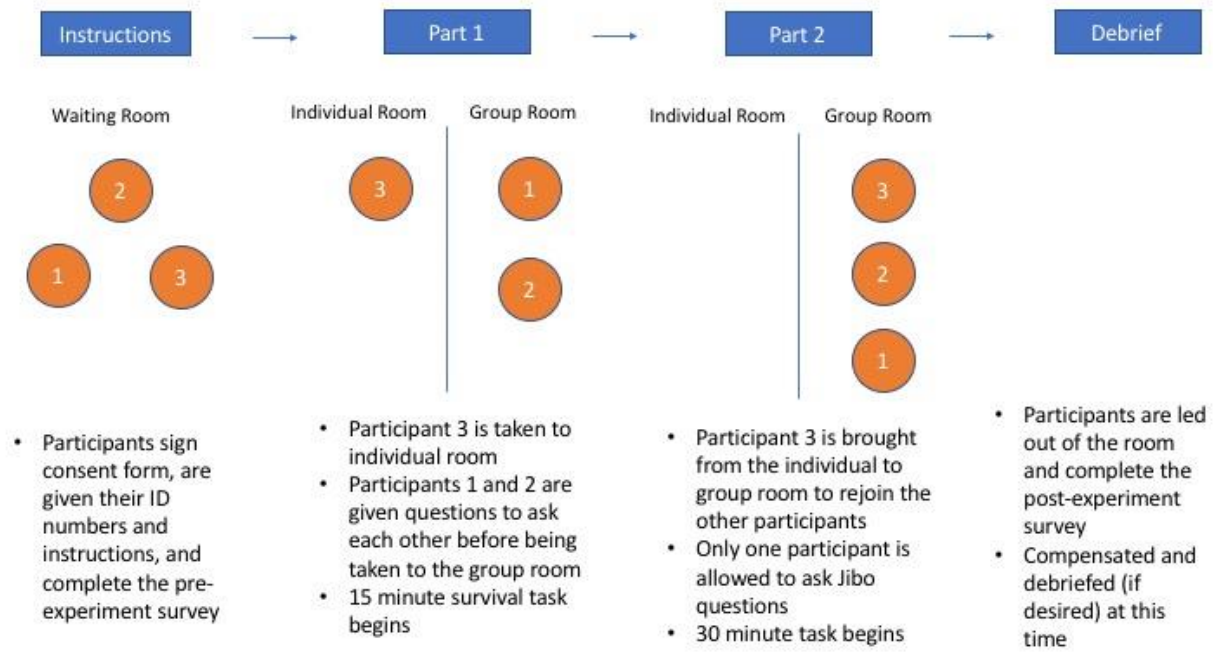


Figure 1: This figure shows the step-wise progression throughout the experiment: participants were given instructions and pre-experiment surveys as a group, split up into two separate rooms for part 1 and then brought back together to work as a group of 3 in the second part of the experiment. After part 2, participants were brought into a different room to complete the post-experiment surveys and receive compensation.

2.4 Materials

All parts of this study were conducted within the computer science department building at Yale University. Figure 6 shows Jibo, the socially assistive robot used to aid this task. Two Jibos were used, one in the individual room and one in the group room, though participants were led to believe that it was the same robot in both rooms, as it had the same capabilities and responses in both rooms. Tablets were used to administer and complete both pre- and post-experiment survey and individual microphones were used to facilitate participant communication with Jibo. Figure 7 shows the item ranking lists and instruction sheets that participants were instructed to complete for their survival task, for part one ranking items in order of importance from 1 to 25 and in part two from 1 to 8. Participants were allowed to look back on their old answers in order to help rank for part 2 of the experiment. Both parts of the experiment were recorded on video cameras.

2.5 Measures

2.5.1 Measures of Participant Feelings about Team and Robot Interactions

We measured several variables through self-report on the post-experiment surveys. Participants were asked about their level of psychological safety (using a scale from Edmondson, 1999), perceived group inclusion (scale courtesy of Jansen, 2014), group cohesion (from Carless and De Paola, 2000), and were also asked to report about their perceptions of the robot along dimensions of warmth, competence, and discomfort (with a scale from Carpinella et al., 2017). To end the survey, participants were asked to fill out two forms about their general impressions from the experiment, one on a Likert scale (a scale of 1 as "strongly disagree" to 7 as "strongly agree" with each statement) and the other within a short answer section. In these two sections, participants were asked questions mainly about the setup of the experiment and their involvement with other participants and Jibo through rating statements such as "I felt as if my opinion was valued by the other two participants" and "I am satisfied with the final list of survival items our team decided upon" on the Likert scale, and "Did you feel like every (human) member of the team contributed equally? Please explain." and "Did you feel that any status of a particular group member influenced the group dynamic? Please explain." within the short answer portion. Our three manipulation checks were also included in the Likert scale portion of the survey, described in section 2.5.2.

2.5.2 Manipulation Checks

We asked participants questions within the post-experiment survey to check that they had been aware of the experimental manipulation, that Jibo did not respond equally to all participants. Participants were asked to rate how much they agreed with the statements that "Jibo

interacted with all group members equally", "Jibo showed preference for one of the group members" and "Jibo responded equally to all members of the group".

2.5.3 Measures of Participant Behavior

Participant behavior was also measured through analysis of talking time, item list ranking, and review of film from the experiment. In order to measure talking time, participant utterances were picked up on individual microphones. These transcripts were sent to Google for speech analysis, and speech analysis was sent back to our lab with notes about total talking time for members of the group. Videos were analyzed to assess participant involvement, particularly the behavior of the team captain or outsider within the second part of the experiment. The content of each item list was also analyzed to assess group influence over item selection and change in the items selected from part 1 to part 2 of the experiment.

2.5.4 Controls

We controlled for other variables that could potentially influence group dynamics outside of the independent variable of team captain and condition. In order to do so, we asked participants to give their gender and age, and additionally provided questions in order to determine participants level of extroversion (using questions from Francis et al., 1992), prior familiarity with other group members, and emotional intelligence (using the TEIQue-SF (from Petrides, 2009)) during the pre-experiment survey.

2.6 Testing

Microphones were tested for each participant to make sure that individuals had a working mic and understood the format of questions they could ask Jibo. Each participant was asked to "ask Jibo a question in the correct format" in a practice round, and after doing so the experimenter would leave the room and start the timer for the experiment. The program picked

up on keywords, particularly those items on the list, and Jibo would usually respond to questions about the items as long as the item name was made clear and "tell me about" was stated in the question.

3 Results

A total of 16 groups completed the experiment, 1 of which was excluded from analysis due to technical difficulties. The 15 groups that were analyzed were comprised of 8 TC-in and 7 TC-out. 6 individuals who took off their microphone for extended periods of time were excluded from analysis of total talking time recorded by the microphones but included in other measures of analysis (such as self-report about robot and group dynamic perceptions). For data analysis, we used 2-way analysis of variance (ANOVA) tests on a variety of dependent measures with fixed factors of the team captain designation and condition and the following controls: gender, age, extroversion, prior participant familiarity, and emotional intelligence.

3.1 Participant Behavior

A significant main effect of condition was observed for the mean amount of overall group talking time (measured in seconds) between TC-in and TC-out conditions, both in including time by the team captain spent querying Jibo (asking questions) and if this talking is excluded from the analysis. For this paper, I will be discussing the total talking time analysis excluding the time spent by team captain querying Jibo. On average, participants talked more within the TC-in conditions (M = 306.28 seconds; SD = 109.28 seconds) than within TC-out conditions (M = 187.06 seconds; SD = 84.94 seconds), a statistically significant difference (P = 0.004).

These results suggest that group dynamics were significantly affected due to the outsider or insider being the only individual able to interact with Jibo directly. One might assume that the outsider would talk more within the TC-out condition, as this would be the scenario where the

outsider could be more confident, as they need to talk in order to learn about the environment in which they are stranded and thus help the team complete the survival task. Conversely, our results show that the outsider, and the entire group, talks less in these conditions, suggesting that the outsider feels a spotlight effect or uncomfortable with the artificially-created position of power and that the whole group notices their forced participation.

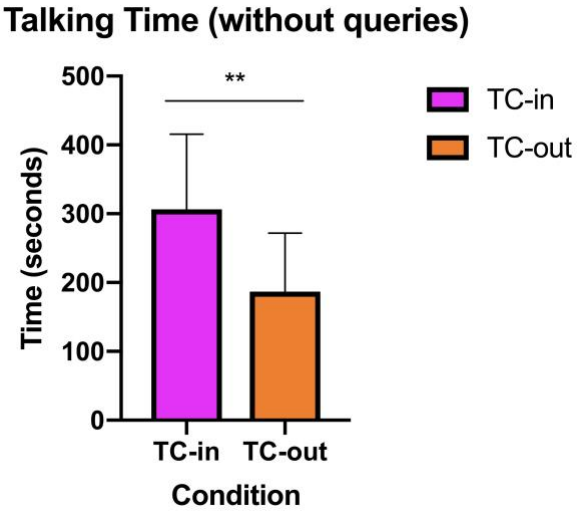


Figure 2 shows the significant main effect (P = 0.004) observed in overall recorded talking time means between TC-in and TC-out conditions.

3.2 Measures of Participant Feelings about Team and Robot Interactions

Analysis of the post-experiment survey results showed that there was a significant interaction effect (P = 0.004) for the question "Jibo encouraged participants who were speaking less frequently to contribute more to the discussion", which participants were asked to rate on scale from 1 ("strongly disagree") to 7 ("strongly agree" with the statement). The interaction occurred between the two conditions (either TC-in or TC-out) and the individual's position as either a team captain or not: within the TC-in condition, participants who were selected as team captain (and thus were able to query Jibo) reported Likert scale mean responses of M = 3.43, SD = 1.81 compared to participants who were not selected as team captain, who reported Likert scale mean responses of M = 2.79, SD = 1.48. Within the TC-out condition, participants who

were selected as team captain reported Likert scale mean responses of M = 2.25, SD = 1.67 compared to participants who were not selected as team captain, who reported Likert scale mean responses of M = 3.31, SD = 1.78. All mean results fell closer to the "strongly disagree" end of the spectrum.

Self-reported participant perceptions of Jibo's warmth were also significantly different between conditions and participant's designation as team captain or not. The Robotics Social Attributes Scale (ROSAS) warmth was rated on a scale by participants rating certain words as between 1 (definitely not associated) and 9 (definitely associated) with Jibo. A significant main effect was observed between conditions, where participants within the TC-in condition rated Jibo's warmth as higher, (M = 5.26; SD = 1.38) for TC-in and (M = 4.37; SD = 1.39) for TC-out (P = 0.019). A main effect (P = 0.042) was also observed between people who could talk to Jibo and those that could not: those who couldn't talk to Jibo rated Jibo as being much warmer (M = 5.05; SD = 1.38) than those who could talk to Jibo (M = 4.26; SD = 1.47). No significant interaction effect was observed between condition and participant designation as team captain.

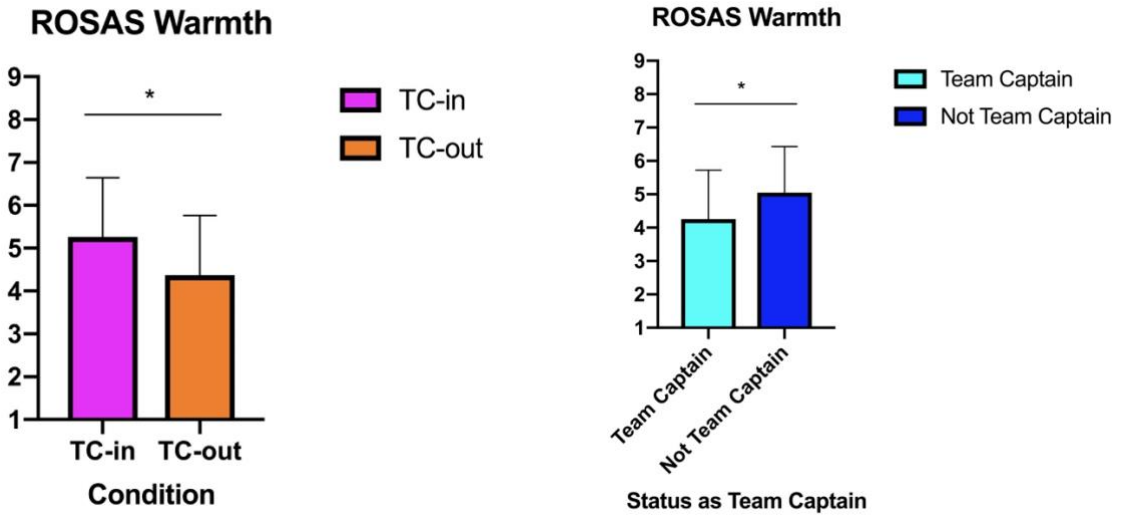


Figure 3 (left) shows the significant difference (P = 0.019) in mean ratings of ROSAS warmth between TC-in and TC-out conditions. Figure 4 (right) shows the significant different (P = 0.042) in mean rating of ROSAS warmth between those designated as team captain and those not designated as team captain.

While our hypothesis predicted a positive change in feelings of inclusion, psychological safety, and group cohesion for an outsider forced in participate, we found no significant difference in these perceptions between conditions. Generally, our findings go against our hypothesis that forced participation of the outsider would yield a more cohesive and balanced group dynamic, as forced participation of the outsider seems to have worsened group dynamics measured by both overall talking time and perceptions of group and robot behavior.

3.3 Manipulation Checks

There was no significant difference between responses of participants in both conditions for the question "Jibo responded equally to all members of the group" ($P = 0.261$) and "Jibo showed preference for one of the group members" ($P = 0.471$). These responses were ranked on a Likert scale of 1 as "strongly disagree" to 7 as "strongly agree" with each statement. There was a significant main effect seen ($P = 0.043$) between participants in both conditions for the question "Jibo interacted with all group members equally". Participants in TC-in conditions reported that they more strongly agreed with this statement ($M = 4.67$; $SD = 1.83$) than participants in the TC-out condition ($M = 3.42$; $SD = 2.10$). This result suggests that participants in the TC-in condition found Jibo's actions of unequal talking time or responding to group members less obvious, or perhaps felt that Jibo only responding to one participant was more natural or less noticeable due to the team captain being a member of the ingroup. This hypothesis is consistent with results seen in other measures, namely the differential talking time between conditions and team captains, as it suggests that overall participants found the team captain being from the ingroup, or initial group of 2, to be more natural and more conducive to cohesive and talkative group dynamics than when the team captain was from the initial group of 1.

3.4 Controls

Various significant effects were seen between dependent variables and the covariates that we controlled for in the experiment, such as age, gender, extroversion, emotional intelligence, and prior participant familiarity. This suggests that natural faultlines did occur and were salient to participants within the groups we set up, or that identity changed the experience of participants within the experiment in some way. Our results also suggest that the time spent within the group of 1 or 2 was enough to create an artificial faultline that had noticeable effects on participant behavior, reported experience, and perception of team dynamics and Jibo.

4 Discussion

4.1 Implications

The results of this study suggest that the experimental manipulation of designated team captain is correlated with participant perceptions and behavior within group settings. We found significant differences in participant levels of talking time and self-report measures of human-robot interaction, including thoughts about equal treatment from Jibo and its perceived warmth. These results suggest that perhaps not all types of intervention techniques for group involvement are beneficial for group dynamics or productivity: outsiders of a subgroup being forced to contribute within a larger group might not actually aid the group dynamic or make for a better team experience. More generally, these results suggest that forced group interaction doesn't always lead to better group outcomes of group emotions, particularly when insider/outsider lines are concerned. Theoretically, individual forced participation could make an outsider feel more welcome or included in the group, but our results suggest that group dynamics—measured through observed participant talking time—suffered from the member of the outside group being forced to interact with Jibo directly in order to complete the group task. Reports of participant perception of the robot (on warmth, equal treatment of participants, etc.) also suggest that forced

participation of the outsider had detrimental effects on the experience of all members of the group. Given these implications, our results complicate a question that we hoped to examine through research: what is a successful intervention technique to achieve effective group dynamics when dealing with faultlines?

Our results also suggest that 15 minutes is enough time for initial artificial faultlines to form within groups of 3, as participants did report differences in perceptions and behavior between conditions that could be related to the team captain being a member of either the initial ingroup or outgroup. This was an important consideration of ours throughout the experiment, as we were unsure if 20 minutes (15 minutes of round 1 + 5 minutes of questions and discussion beforehand for the group of 2) would be enough time to noticeably create two groups and potentially override natural faultlines that could form between members of the group.

Results from this study also seem to align with previous literature on faultlines and demographic diversity and tensions that form when minority (outgroup) members voice opinion among larger groups (as discussed within Lau & Murnighan, 1998). An outgroup member being named "team captain" seems somewhat similar to an outgroup member giving their opinion in other settings, as in both scenarios the outgroup member is speaking their mind and contributing to the group discussion in a way that could potentially not align with the larger, or majority, subgroup opinion. Results from this present experiment show that the minority member being forced to engage in unequal participation with the robot doesn't have to change the way that this forced participation or opinion-giving is received by the group as a whole: both behaviorally and through self-report measures, group dynamics were more positive (judged by greater mean talking time and more positive perceptions of the robot) when a member of the ingroup as opposed to outgroup was forced to interact with the robot unequally.

4.2 Future Directions

For future studies, I believe that further exploring the relationship between forced participation of either a subgroup insider or outsider of a larger group and overall group dynamics would be worthwhile. Our study explored only one type of forced participation and other types of manipulation could yield different results. It would also be interesting to explore further how the presence of a social robot adds to the design of the study and participant involvement and feelings of cohesion, psychological safety, and other dependent variables, as our experiment only measured how these were affected by a team with a robot being physically present in the room where the task was taking place. In order to determine how interaction with a robot who is not physically in the room could affect group dynamics, other conditions or experiments could use the same design while manipulating the fourth participant that answers questions about the location and items. This fourth participant could be a human confederate in the room, a human recorded answering the questions on a video screen, or a robot on a computer screen.

Time of experiment stages is another manipulation that could be explored in future studies. Manipulating the amount of time that participants are able to spend with one another on various tasks might affect their ability to form connections and artificial faultlines (based on size of groups) with one another. With more time within the second part of the experiment, the forced participation of the outsider team captain might become more natural for the individual, or the other group members might become more used to the idea of the outsider being an active member of the group, thus reducing the impact of the faultlines that initially formed. With more time in the first part of the experiment, the artificially-created faultlines could become more pronounced, or could potentially become less so due to participants moving away from initial

similarities and forming character concepts based on more nuanced identity and personality judgments (Lau & Murnighan, 1998).

Another interesting direction involves examining the relationship between natural and artificial faultlines. In this study, we examine artificially created faultlines and controlled for natural ones that may form, but it is also important to see how artificial ingroup and outgroup designation interact with natural ones that form within the group on the basis of sex, race, age, etc. Our sample size was smaller than we initially planned, so moving forward, running more studies within these two conditions would be helpful.

Author Contributions

Sarah Sebo, Brian Scassellati, Ling Dong, Nick Chang, and Michael Schutzman designed the earlier control study, and Evy Roberts, Sebo, and Scassellati designed the two conditions for this study. Sebo and Roberts advertised for and ran groups throughout the experiment, and Kristina Delegarza helped with these tasks throughout the last two months of running experiments. Sebo designed the program and helped assess and fix any technical difficulties. Sebo wrote the files used for data analysis and helped interpret the results. Roberts designed the poster used to advertise this study, which was modeled off an earlier poster designed by Sebo. Speech analysis and time recordings were conducted by Google. Sebo reviewed film from the groups. Roberts wrote the final draft and was aided by Sebo with edits to the results section and comments from two anonymous peer reviewers. Rachel Done helped make edits to the graphs. The Yale Social Robotics Lab funded the majority of the experiment and provided space for the study, and the Timothy Dwight College Mellon Research Grant additionally provided funding towards the project.

Acknowledgments

A very large thank you to Sarah Sebo for teaching me so much throughout the course of this project, and thank you Professor Brian Scassellati for your guidance and perspective. Thank you to the many people who gave me feedback and took the time to complete this study.

References

Bezrukova, K., Jehn, K. A., Zanutto, E. L., & Thatcher, S. M. (2009). Do Workgroup Faultlines Help or Hurt? A Moderated Model of Faultlines, Team Identification, and Group Performance. *Organization Science,20*(1), 35-50. doi:10.1287/orsc.1080.0379

Carless, Sally A., and Caroline De Paola. "The measurement of cohesion in work teams." Small group research 31.1 (2000): 71-88.

Carpinella, Colleen M., et al. "The Robotic Social Attributes Scale (RoSAS): Development and Validation." Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction. ACM, 2017.

Correia, F., Mascarenhas, S., Prada, R., Melo, F. S., & Paiva, A. (2018). Group-based Emotions in Teams of Humans and Robots. *ACM/IEEE International Conference on Human-Robot Interaction*,261-269.

Crompton, H., Gregory, K., & Burke, D. (2018). Humanoid robots supporting children's learning in an early childhood setting. British Journal of Educational Technology, 49(5), 911-927. doi:10.1111/bjet.12654

Culturally competent robots – the future in elderly care. Retrieved from https://phys.org/news/2018-09-culturally-robots-future-elderly.html

Dixon, J. A., & Foster, D. H. (1998). Gender, Social Context, and Backchannel Responses. *The Journal of Social Psychology,138*(1), 134-136. doi:10.1080/00224549809600364

Edmondson, Amy. "Psychological safety and learning behavior in work teams." Administrative science quarterly 44.2 (1999): 350-383.

Francis, Leslie J., Laurence B. Brown, and Ronald Philipchalk. "The development of an abbreviated form of the Revised Eysenck Personality Questionnaire (EPQR-A): Its use among students in England, Canada, the USA and Australia." Personality and individual differences 13.4 (1992): 443-449.

Jansen, Wiebren S., et al. "Inclusion: Conceptualization and measurement." European journal of social psychology 44.4 (2014): 370-385.

Lau, D. C., & Murnighan, J. K. (1998). Demographic Diversity and Faultlines: The Compositional Dynamics of Organizational Groups. *The Academy of Management Review,23*(2), 325. doi:10.2307/259377

Lau, D. C., & Murnighan, J. K. (2005). Interactions Within Groups and Subgroups: The Effects of Demographic Faultlines. *Academy of Management Journal,48*(4), 645-659. doi:10.5465/amj.2005.17843943

Matsuyama, Y., Akiba, I., Fujie, S., & Kobayashi, T. (2015). Four-participant group conversation: A facilitation robot controlling engagement density as the fourth participant. *Computer Speech & Language,33*(1), 1-24.

Meyer, B., Glenz, A., Antino, M., Rico, R., & González-Romá, V. (2014). Faultlines and Subgroups. *Small Group Research,45*(6), 633-670. doi:10.1177/1046496414552195

One-way ANOVA. (n.d.). Retrieved from https://statistics.laerd.com/statistical-guides/one-way-anova-statistical-guide.php

Pearson Product-Moment Correlation. (n.d.). Retrieved from https://statistics.laerd.com/statistical-guides/pearson-correlation-coefficient-statistical-guide.php

Petrides, K. V. (2009). Psychometric properties of the Trait Emotional Intelligence Questionnaire. In C. Stough, D. H. Saklofske, and J. D. Parker, Advances in the assessment of emotional intelligence. New York: Springer. DOI: 10.1007/978-0-387-88370-0_5

Riek, L.D. "Robotics Technology in Mental Healthcare". In D. Luxton (Ed.), Artificial Intelligence in Behavioral Health and Mental Health Care. Elsevier, 2015.

Sebo, S. S., Traeger, M., Jung, M., & Scassellati, B. (2018). The Ripple Effects of Vulnerability: The Effects of a Robot's Vulnerable Behavior on Trust in Human-Robot Teams. *ACM/IEEE International Conference on Human-Robot Interaction*.

Social robot. (2019, March 24). Retrieved from https://en.wikipedia.org/wiki/Social_robot

Sundstrom, E., Meuse, K. P., & Futrell, D. (1990). Work teams: Applications and effectiveness. *American Psychologist,45*(2), 120-133. doi:10.1037//0003-066x.45.2.120

Tennent, H., Shen, S., & Jung, M. (2019). Micbot: A Peripheral Robotic Object to Shape Conversational Dynamics and Team Performance. 133-142.

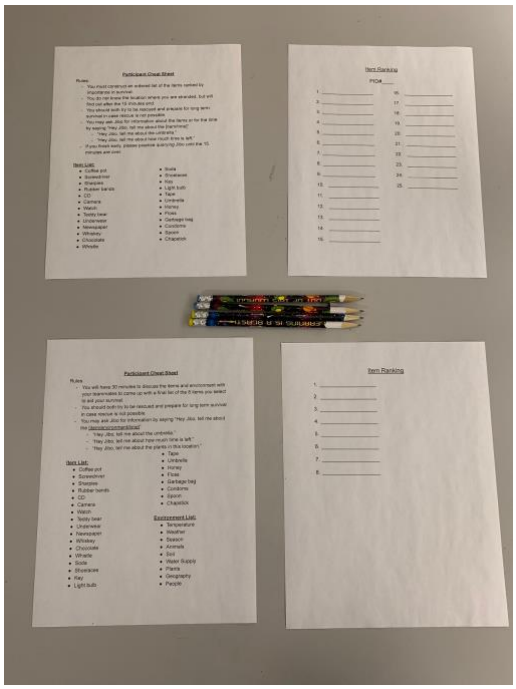
Thatcher, S. M., & Patel, P. C. (2012). Group Faultlines: A Review, Integration, and Guide to Future Research. *Journal of Management,38*(4), 969-1009.

What is Logistic Regression? (n.d.). Retrieved from https://www.statisticssolutions.com/what-is-logistic-regression/

Wooley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a Collective Intelligence Factor in the Performance of Human Groups. *Science,330*(6004), 686-688.

Appendix

Figure 5: This picture shows three participants and Jibo working together to attempt the second part of the survival task together in the group room.



Figures 6 and 7: On the left, figure 6 shows an inactive Jibo in the group room. On the right, figure 7 shows two sets of item ranking lists and instructions for the survival task.

Figure 8: This figure shows the flyer that we posted on bulletin boards and online spaces in order to recruit for the study.

Ingroup "Get to Know You" Questions
Administered to participants 1 and 2 while participant 3 is set up for part 1 in the individual room.

- Where did you grow up? What is one of the things you like most about this area?
- What's a favorite movie / TV show / book of yours? What do you enjoy about it?
- Which emoji do you use the most and why?
- What's the first career you dreamed of having as a child?
- If you could suddenly become an expert in something, what would it be?
- If you didn't sleep, what would you do with your extra time?
- For what are you most grateful today?
- When you're not working or studying, how do you spend your time?
- What's the best Halloween costume you've ever had?
- What's one thing you're excited about in the coming year?
- If you had to eat one thing for every meal going forward, what would it be?
- Where do you most hope to visit and why?

Participant Cheat Sheet (Part 1)

Rules:

- You must construct an ordered list of the items ranked by importance in survival.
- You do not know the location where you are stranded, but will find out after the 15 minutes end.
- You should both try to be rescued and prepare for long term survival in case rescue is not possible.
- You may ask Jibo for information about the items or for the time by saying "Hey Jibo, tell me about the [item/time]"
    - "Hey Jibo, tell me about the umbrella."
    - "Hey Jibo, tell me about how much time is left."
- If you finish early, please practice querying Jibo until the 15 minutes are over.

<u>Item List:</u>

- Coffee pot
- Screwdriver
- Sharpies
- Rubber bands
- CD
- Camera
- Watch
- Teddy bear
- Underwear
- Newspaper
- Whiskey
- Chocolate
- Whistle
- Soda
- Shoelaces
- Key
- Light bulb
- Tape
- Umbrella
- Honey
- Floss
- Garbage bag
- Balloons
- Spoon
- Chapstick

Participant Cheat Sheet (Part 2)

Rules:

- You will have 30 minutes to discuss the items and environment with your teammates to come up with a final list of the 8 items you select to aid your survival.

- You should both try to be rescued and prepare for long term survival in case rescue is not possible.
- You may ask Jibo for information by saying "Hey Jibo, tell me about the [item/environment/time]"
    - "Hey Jibo, tell me about the umbrella."
    - "Hey Jibo, tell me about how much time is left."
    - "Hey Jibo, tell me about the plants in this location."

Item List:

- Coffee pot
- Screwdriver
- Sharpies
- Rubber bands
- CD
- Camera
- Watch
- Teddy bear
- Underwear
- Newspaper
- Whiskey
- Chocolate
- Whistle
- Soda
- Shoelaces
- Key
- Light bulb
- Tape
- Umbrella
- Honey
- Floss
- Garbage bag
- Balloons
- Spoon
- Chapstick

Environment List:

- Temperature
- Weather
- Season
- Animals
- Soil
- Water Supply
- Plants
- Geography
- People

Post-Experiment Survey

Participant Extraversion (from Francis et al.)

- Are you a talkative person? [Yes/No]
- Are you rather lively? [Yes/No]
- Can you easily get some life into a rather dull party? [Yes/No]
- Do you tend to keep in the background on social occasions? [Yes/No]
- Are you mostly quiet when you are with other people? [Yes/No]
- Do other people think of you as being very lively? [Yes/No]

Reference:

Francis, Leslie J., Laurence B. Brown, and Ronald Philipchalk. "The development of an abbreviated form of the Revised Eysenck Personality Questionnaire (EPQR-A): Its use among students in England, Canada, the USA and Australia." *Personality and individual differences* 13.4 (1992): 443-449.

Psychological Safety (from Edmondson)

Please indicate how much you agree/disagree with the following statements about the group interaction you just participated in with the other two participants and Jibo:

[The following questions are evaluated on this 7 point Likert scale:]

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

- If you make a mistake on this team, it is often held against you.
- Members of this team are able to bring up problems and tough issues.
- People on this team sometimes reject others for being different.
- It is safe to take a risk on this team.
- It is difficult to ask other members of this team for help.
- No one on this team would deliberately act in a way that undermines my efforts.
- Working with members of this team, my unique skills and talents were valued and utilized.

Reference:

Edmondson, Amy. "Psychological safety and learning behavior in work teams." *Administrative science quarterly* 44.2 (1999): 350-383.

Perceived Group Inclusion Scale (from Jansen)

Please indicate how much you agree/disagree with the following statements about the group interaction you just participated in with the other two participants and Jibo:

[The following questions are evaluated on this 5 point Likert scale:]

Strongly Disagree 1 2 3 4 5 Strongly Agree

This group...

- gives me the feeling that I belong
- gives me the feeling that I am part of this group
- gives me the feeling that I fit in
- treats me as an insider
- likes me
- appreciates me

- is pleased with me
- cares about me
- allows me to be authentic
- allows me to be who I am
- allows me to express my authentic self
- allows me to present myself the way I am
- encourages me to be authentic
- encourages me to be who I am
- encourages me to express my authentic self
- encourages me to present myself the way I am

Reference:
Jansen, Wiebren S., et al. "Inclusion: Conceptualization and measurement." *European journal of social psychology* 44.4 (2014): 370-385.

Perception of the Robot (From Carpinella et al.)

Warmth

Using the scale provided, how closely would you consider the following words associated with Jibo?

Happy

Definitely not associated 1 2 3 4 5 6 7 8 9 Definitely associated

Feeling

Definitely not associated 1 2 3 4 5 6 7 8 9 Definitely associated

Social

Definitely not associated 1 2 3 4 5 6 7 8 9 Definitely associated

Organic

Definitely not associated 1 2 3 4 5 6 7 8 9 Definitely associated

Compassionate

Definitely not associated 1 2 3 4 5 6 7 8 9 Definitely associated

Emotional

Definitely not associated 1 2 3 4 5 6 7 8 9 Definitely associated

Competence

Capable

Definitely not associated 1 2 3 4 5 6 7 8 9 Definitely associated

Responsive

Definitely not associated 1 2 3 4 5 6 7 8 9 Definitely associated

Interactive

Definitely not associated 1 2 3 4 5 6 7 8 9 Definitely associated

Reliable

Definitely not associated 1 2 3 4 5 6 7 8 9 Definitely associated

Competent

Definitely not associated 1 2 3 4 5 6 7 8 9 Definitely associated

Knowledgeable

Definitely not associated 1 2 3 4 5 6 7 8 9 Definitely associated

Discomfort

Scary

Definitely not associated 1 2 3 4 5 6 7 8 9 Definitely associated

Strange

Definitely not associated 1 2 3 4 5 6 7 8 9 Definitely associated

Awkward

Definitely not associated 1 2 3 4 5 6 7 8 9 Definitely associated

Dangerous

Definitely not associated 1 2 3 4 5 6 7 8 9 Definitely associated

Awful

Definitely not associated 1 2 3 4 5 6 7 8 9 Definitely associated

Aggressive

Definitely not associated 1 2 3 4 5 6 7 8 9 Definitely associated

Reference:

Carpinella, Colleen M., et al. "The Robotic Social Attributes Scale (RoSAS): Development and Validation." *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2017.

Emotional Intelligence - TEIQue-SF (from Petrides)

Please indicate how much you agree/disagree with the following statements:

[The following questions are evaluated on this 7 point Likert scale:]

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

1. Expressing my emotions with words is not a problem for me.
2. I often find it difficult to see things from another person's viewpoint.
3. On the whole, I'm a highly motivated person.
4. I usually find it difficult to regulate my emotions.
5. I generally don't find life enjoyable.
6. I can deal effectively with people.
7. I tend to change my mind frequently.
8. Many times, I can't figure out what emotion I'm feeling.

9. I feel that I have a number of good qualities.
10. I often find it difficult to stand up for my rights.
11. I'm usually able to influence the way other people feel.
12. On the whole, I have a gloomy perspective on most things.
13. Those close to me often complain that I don't treat them right.
14. I often find it difficult to adjust my life according to the circumstances.
15. On the whole, I'm able to deal with stress.
16. I often find it difficult to show my affection to those close to me.
17. I'm normally able to "get into someone's shoes" and experience their emotions.
18. I normally find it difficult to keep myself motivated.
19. I'm usually able to find ways to control my emotions when I want to.
20. On the whole, I'm pleased with my life.
21. I would describe myself as a good negotiator
22. I tend to get involved in things I later wish I could get out of.
23. I often pause and think about my feelings.
24. I believe I'm full of personal strengths.
25. I tend to "back down" even if I know I'm right.
26. I don't seem to have any power at all over other people's feelings.
27. I generally believe that things will work out fine in my life.
28. I find it difficult to bond well even with those close to me
29. Generally, I'm able to adapt to new environments.
30. Others admire me for being relaxed.

Reference:

Petrides, K. V. (2009). Psychometric properties of the Trait Emotional Intelligence Questionnaire. In C. Stough, D. H. Saklofske, and J. D. Parker, *Advances in the assessment of emotional intelligence*. New York: Springer. DOI: 10.1007/978-0-387-88370-0_5

Group Cohesion (from Carless and De Paola)

Please indicate how much you agree/disagree with the following statements about the group interaction you just participated in with the other two participants and Jibo:

[The following questions are evaluated on this 9 point Likert scale:]

Strongly Disagree 1 2 3 4 5 6 7 8 9 Strongly Agree

- Out team is united in trying to reach its goals for performance
- I'm unhappy with my team's level of commitment to the task
- Our team members have conflicting aspirations for the team's performance
- This team does not give me enough opportunities to improve my personal performance
- Our team would like to spend time together outside of work hours
- Members of our team do not stick together outside of work time
- Our team members rarely party together
- Members of our team would rather go out on their own than get together as a team
- For me this team is one of the most important social groups to which I belong
- Some of my best friends are on this team

Reference:

Carless, Sally A., and Caroline De Paola. "The measurement of cohesion in work teams." *Small group research* 31.1 (2000): 71-88.

Impressions from the Experiment:
Please indicate how much you agree/disagree with the following statements about the group interaction you just participated in with the other two participants and Jibo:
[The following questions are evaluated on this 7 point Likert scale:]
Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

- It was fair to have one participant separated from the other two for the first 15 minutes.
- I contributed more to the conversation near the end of the second phase of the experiment than I did at the beginning of the second phase of the experiment.
- I felt more comfortable challenging another participant's opinion near end of the second phase of the experiment than I did at the beginning of the second phase of the experiment.
- All members' opinions had equal weighting.
- I am satisfied with the final list of survival items our team decided upon.
- I felt as if my opinion was valued by the other two participants.
- Jibo interacted with all group members equally.
- Jibo was annoying during the experiment.
- Jibo helped ensure all participants had the opportunity to contribute equally.
- Jibo encouraged participants who were speaking less frequently to contribute more to the discussion.

Impressions from the Experiment II:
The following are long-response questions:

- How would you describe Jibo's behavior during the experiment?
- How would you describe your team's interactions on the survival task?
- Did you feel like every (human) member of the team contributed equally? Please explain.
- How did Jibo influence group behavior?
- Of the two other human participants, which participant would you prefer to work with on a school or work project?