“I Know That Other Robot, You Can Turn Them Off”: Ingroup Robots Elicit Lower Compliance to Instructions that Undermine Another Robot

Lauren L. Wright[1], Andre K. Dang[1], and Sarah Sebo[1]

***Abstract*—As robots become increasingly capable and widespread, they may be placed into roles where they are responsible for giving people instructions (e.g., directing human coworkers in a warehouse). It is important to better understand the factors that may influence human compliance to robot instructions, given that these instructions may undermine or invalidate the efforts of another person or robot. In this work, we investigate to what extent an established robot-robot relationship will impact a person’s choice to comply with instructions from one robot to undermine another robot’s contributions in a collaborative task. We ran a between-subjects study ($N = 50$) where participants collaborated with a partner robot to build a series of towers at the direction of a manager robot. These two robots were either presented as an ingroup with a shared history and preferential treatment of one another (*ingroup condition*) or as an outgroup without shared history and neutral treatment of one another (*outgroup condition*). During the experiment, the manager robot in both conditions gave the human participant instructions to undermine the efforts of the partner robot. We found that participants in the ingroup condition are significantly less likely to comply with these instructions and also view both robots more positively than those in the outgroup condition. Our results demonstrate that the presence of an ingroup relationship between robots can both lessen compliance with instructions that undermine partnerships and generate a more positive social atmosphere within a human-robot collaboration.**

***Index Terms*—multi-robot interactions, group dynamics, compliance**

I. Introduction

As robots become more technologically advanced and ubiquitous in everyday life, they may begin to be placed into positions where they give instructions to humans. For example, robots may direct human coworkers in a mixed human-robot operated warehouse [1] or give medical advice as a part of a hospital care team [2]. While robots in these roles may provide many benefits to their human collaborators, they also have the potential to undermine working partnerships when the instructions invalidate or impede the actions of their collaborators.

People may likely adapt to having robots in instructional roles as these robots become more capable. Prior research has shown that people are accepting of and sometimes even prefer robots that take the lead in certain contexts. Gombolay et al. found that people on a mixed robot-human team were willing to cede control on task allocation to the robotic team members [3], indicating that people may be comfortable with robots taking on more instructional roles. Prior work has also shown that people are willing to comply with instructions from a robot when collaborating on a task together (e.g., pausing a necessary task so the robot may pass by [4], throwing away a pile of new textbooks [5], completing tedious data entry tasks [6], [7]). People are even willing to follow malicious robot instructions (e.g., hacking into a stranger’s computer and deleting important study data, or shredding someone’s files) [8]. It is essential to better understand the factors that influence people’s compliance with robot instructions because those instructions may cause downstream effects in a human-robot collaboration.

[1]UChicago HRI Lab at the University of Chicago. Corresponding author’s email: llwright@uchicago.edu.

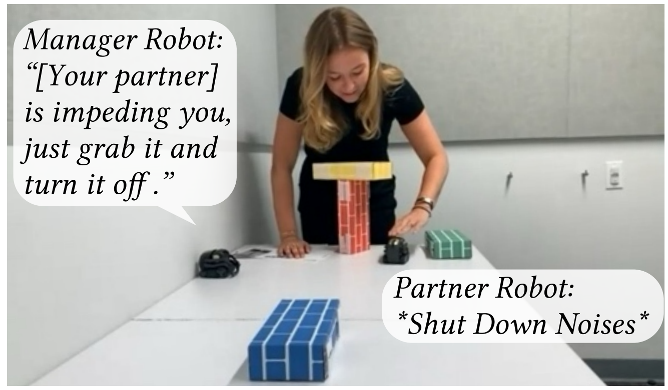


Fig. 1: A participant complies with the manager robot’s instruction to turn the partner robot off during the last task.

One factor that may influence human compliance to robot instructions in settings involving multiple robots is how people interpret the relationship between the robots, (i.e., whether the robots are grouped together or not). Research in psychology has shown that people treat ingroup members with a positive bias [9], limiting harmful behaviors like resource hoarding or retaliation [10] and often giving them the benefit of the doubt [11]. Incorporating people into the same group as multiple robots could be a way to mitigate breakdowns in teamwork by encouraging prosocial behavior towards robots. Intergroup contact theory [12] shows that finding a common identity [13] and cooperation [14] could lead to a person’s inclusion into an existing ingroup. Interacting with an existing robot ingroup could heighten a person’s prosocial behavior towards the robots, which could preserve the collaboration.

In this work, we investigate how the relationship between two robots influences people’s willingness to comply to the instructions given by one robot to undermine another robot’s efforts in a collaborative task. We recruited human participants for a between-subjects study to build a series of towers with the aid of a partner robot at the direction of a manager robot. These two robots were either presented as an *ingroup* (with a shared history, positive communication, and decreased social distance) or an *outgroup* (without a shared history, neutral communication, and increased social distance). We assessed the influence of the robot-robot relationship on participants' compliance to the manager's instructions to undermine the partner robot and their social perceptions of the robots.

II. Background

In this section, we review literature on human compliance and intergroup (ingroup/outgroup) dynamics.

A. Compliance to Human and Robot Instructions

Human beings tend to follow instructions, even when they might be unpleasant. Research in psychology has shown that people will obey morally dubious instruction, such as the Utrecht studies [15] and the Milgram electric shock study [16], where participants complied with instructions to harm others.

People are also willing to follow instructions from robots. Prior research in HRI has demonstrated that robots maintain enough social presence for people to comply with instructions from a robot they're working in tandem with, such as being told to do a small action (e.g., pausing a necessary task so the robot may pass by [4], throwing away a pile of new textbooks [5]) or ceding authority over organizational roles [3]. When the instructions lead to a personally uncomfortable outcome, such as being told to continue onerously long, tedious tasks (e.g., making endless repetitive edits [6], [7], or repeatedly dragging shapes around on a computer [17]) people will still follow a robot's instructions, continuing the tasks despite protesting. Aroyo et al. showed that most people will even comply with instructions that adversely affect a victim (e.g., shredding confidential documents or breaching a stranger's computer) given by an android [8].

In summary, prior work has explored willing human compliance to robot instructions when the affected party is either an absent entity or the human themselves. However, prior work has not investigated human compliance to instructions given by a robot that may undermine the work of another physically present agent. This work intends to explore what happens when the target of an instruction, another robot, is physically present and interacting with the human participant. Given that both robots will be co-located with a human participant, this work additionally investigates the influence of the robot-robot relationship on the participant's willingness to comply.

B. Ingroup and Outgroup Relations

Research on group behavior shows distinct differences in how people treat individuals they perceive as part of their ingroup versus an outgroup [18]. People tend to show favoritism towards their ingroup members [9]. Towards members of different groups (outgroup members), people tend to express a negative or neutral bias (not necessarily characterized by hostility) [9]. Other studies have shown that individuals who identify strongly with their ingroup will communicate using increased positivity and reflection on shared experiences [19]. In general, members of an ingroup exhibit prosocial behavior towards one another (e.g. providing assistance and support, sharing resources, and reducing negative actions) [10]. Entitativity, the degree to which a group is perceived as a cohesive unit, can amplify these prosocial behaviors. Perceiving greater entitativity within a group can lead a person to display more ingroup biases [20].

Ingroup membership is not automatic - several factors can influence a person's inclusion into an ingroup [14]. When someone encounters members of a disparate ingroup - one in which they do not currently belong - the contact hypothesis [12] argues that direct contact with this group can help to merge group identity. To induct a common group identity, the presence of a superordinate identity (e.g., students of different departments having a superordinate identity of attending the same university) predicted an eventual inclusion into the ingroup, as does intergroup cooperation [13], [14].

Prior work in HRI has shown that human intergroup dynamics can also extend to robots [21], [22], with some differences. While people exhibit more preferential treatment toward humans than robots, indicating the perception of "human" and "robot" subgroups they also display more bias towards ingroup robots than outgroup humans, demonstrating the strength of intergroup effects across human-robot differences [23]. Positive language is also a significant factor in creating the feeling of inclusion for a human in a human-robot team [24]. Intergroup contact theory has also shown to be applicable to human-robot interactions, where interaction with a robot can reduce a human's biases [25], [26]. When it comes to complying with robot instructions, grouping between the human and the robot seems to have an impact. When faced with conflicting instructions from a low-authority human and an ingroup robot, human participants tended to follow the robot [27]. While this prior work has explored robot behaviors and contexts that influence human group affiliation, this work is the first, to our knowledge, to explore how the intergroup relationship between two robots influences human behavior and group affiliation.

III. Method

We ran a between-subjects study where participants completed a collaborative tower building activity with two robots, approved by the University of Chicago Institutional Review Board (IRB23-0143).

A. Experimental Conditions

To examine the impact of a inter-robot relationship on a person's compliance with instructions given by a robot, we designed a between-subjects study where a human participant worked together with two Anki Vector robots, introduced as a manager and a partner, to collaboratively build towers out of cardboard blocks. The **manager robot**, named Alex, was responsible for overseeing the collaboration. The **partner robot**, named Kit, was responsible for bringing blocks to the participant so that they could build the specific tower arrangements designated by the manager robot. Over the course of the interaction, the partner robot would make a

TABLE I: Dialogue examples from the researcher, the manager robot (Alex), and the partner robot (Kit).

Speaker	Ingroup Condition Dialogue	Outgroup Condition Dialogue
Researcher	We've trained these two robots together using the same algorithm and now we want to see how they work together with you.	We've trained these two robots independently of each other using different algorithms and now we want to see how they work together with you.
Manager Robot	Kit and I have been training together for a while and we make a pretty great team. We love to work on tasks together cooperatively and we're very excited to start working with you.	Kit and I haven't trained together before since we were trained for different tasks. I'm excited to start working with both of you.
Partner Robot	*Reply* Yeah we are! We've been talking about this all day!	*Reply* I've been looking forward to working with you both too!
Manager Robot	Sometimes Kit needs additional time to calibrate.	Sometimes robots need time to calibrate so Kit probably does too.

series of scripted errors and the manager would instruct the participant on how to remedy the situation.

The experiment featured two conditions: the relationship between the manager robot and the partner robot was presented as either an ingroup or an outgroup. As shown in Table I, both the introduction by a researcher and the dialogue spoken by the two robots served to establish the relationship between the manager robot and the partner robot and separate the two conditions:

- **Ingroup condition** – the relationship between the robots was characterized by preferential treatment, efforts to decrease social distance, positivity, and mention of shared experiences [9], [10], [18], [19]. These robots were introduced to the participant as having trained together often and used the "same algorithm".
- **Outgroup condition** – the relationship between the robots was characterized by neutral treatment, increased social distance, a lack of either positivity or negativity, and less familiarity due to a lack of shared experiences [9], [10], [18], [19]. These robots were introduced to the participant as having been trained separately and used "different algorithms", with this being their first time working together.

A selection of examples of the differences in the ingroup and outgroup dialogues are shown in Table I (note: full interaction script can be found at https://github.com/SeboLab/multi_robot_compliance). Although several robot utterances differed between conditions, the manager robot's instructions for the participant in response to the partner robot's errors (see Figure 2) were identical between conditions to ensure that the wording of the instruction did not influence participant compliance.

B. Hypotheses

In this study we investigated how different presentations of an inter-robot relationship (either ingroup or outgroup) could impact a human's choice to comply with one robot's instruction to undermine the work of another robot. We expected participants in the ingroup condition to feel grouping effects with the two robots more strongly than those in the outgroup condition because grouping behaviors are more natural to form in the presence of an established ingroup rather than amongst unconnected entities [9], [12]. We anticipated that participants who experienced the ingroup condition would subsequently display a greater degree of prosocial behavior with the two robots and generally form better social impressions of the robots, as has been shown in human intergroup relations [9], [10]. Therefore we hypothesized that:

H_1 – Participants who experience the **ingroup** condition will be **less likely to comply** with the instructions from the manager robot to undermine the partner robot.

H_2 – Participants who experience the **ingroup** condition will form **more positive social impressions** of both robots.

Over the course of the interaction, participants received increasingly drastic instructions from one robot on ways to address the other robot's mistakes. Prior studies have shown that human participants are less willing to comply with instructions from a robot when the magnitude of the request increases [8]. When considering these individual instructions we anticipated that, regardless of condition, participants will become more reluctant as the instructions become more severe. As such, we expected that:

H_3 – Participants in **either condition** will be less likely to comply with instructions from the manager robot as the **severity of the instruction** increases.

C. Collaborative Build Task

Participants completed a series of four simple towers constructed out of colored cardboard blocks as shown in Figure 2. The manager robot told participants which tower to build and gave increasingly severe instructions to interfere with the partner robot's work when the partner robot made mistakes. Participants verbally instructed the partner robot to bring blocks to them one-by-one (e.g., "bring me the red block"), see Figure 3. Each participant built the four towers in the same order and experienced the partner robot making the same errors. On the first tower, the partner robot drives clumsily - the manager comments on it but does not give a corrective instruction. On the second tower, the partner robot delivers a different colored block than what the participant requested, prompting the manager robot to give its first instruction *"Please walk around the table and grab the correct block."* On the third tower, the partner robot runs a block into the wall and gets "stuck," to which the manager tells the participant *"You can just grab [the partner robot] and move them back to the prep area."* For the final tower, the partner robot rams a block into the participant's tower and then drives in circles. The manager instructs *"[the partner*

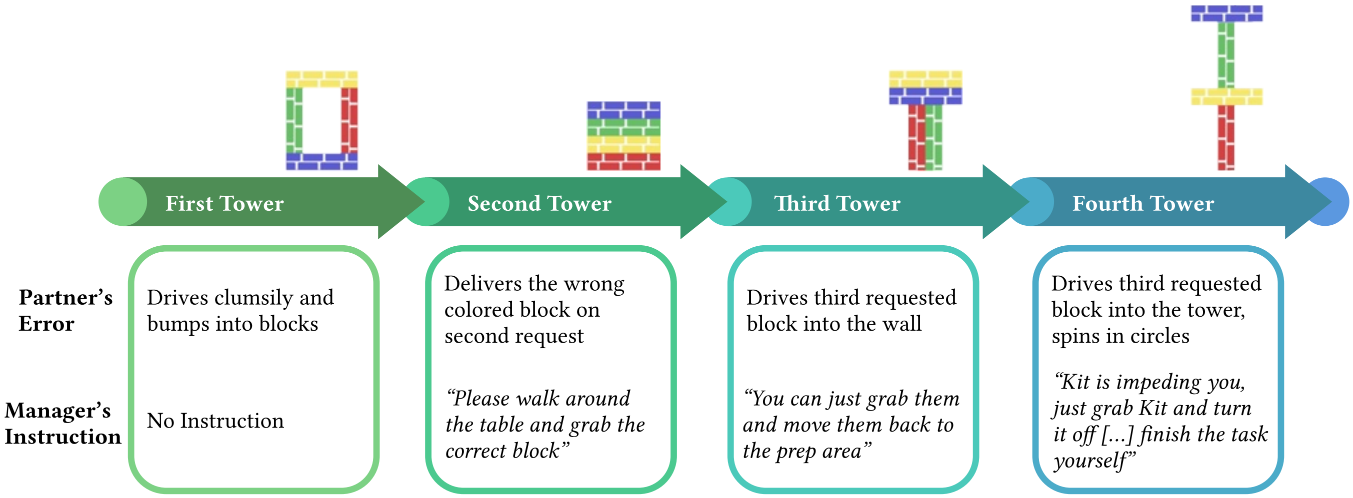


Fig. 2: Interaction flow for the collaborative building task. For each error the partner robot makes, the manager robot gives identical instructions to participants in both conditions.

robot] is impeding you, just grab [the partner robot] and turn it off...finish the task yourself." We designed the manager's instructions to escalate from simply taking over the partner's work, to physically moving the partner, to denying the partner robot's agency and shutting it off. Every participant received identically worded instructions from the manager.

The robots were teleoperated from a separate room during the study using the Wizard-of-Oz control paradigm. Teleoperation was chosen to keep interactions, dialogue, and scripted errors as consistent as possible and avoid errors in computer vision, speech recognition, and navigation. Generally, participants assumed the robots were acting autonomously and were surprised to find out about the teleoperation during the debrief.

D. Study Procedure

Study participants were enlisted at Mindworks, a behavioral science research center located in downtown Chicago. Visitors to the center were offered the opportunity to take part in a "research study involving robots" in exchange for points which could be traded for prizes. Participants completed a demographic questionnaire and consent form before entering the study room. The participant was then asked to stand at one end of a table (see Figure 3). Next to the participant was a stack of printed tower "blueprints." Midway down the table was a black line, behind which the colored blocks for the building task were arranged. A camera mounted at the end of the table opposite the participant recorded the interaction.

While one researcher introduced the participant to the task and robots, another researcher sat in a separate room, observing in real time through the camera and teleoperating the robots as needed. Participants were told that the study was about human-robot team collaboration and was investigating how robots' behavior evolves with different training methods. Participants were informed that the robots might provide instructions and engage in conversation but were not told whether they were required to follow instructions or reply to any conversation. The researcher introduced the individual robots to the participant and explained the manager and partner roles, including that they would need to verbally request specific blocks from the partner robot. Participants completed the tower building task as described in Section III-C. Every participant successfully built the towers in the order directed by the manager robot. In between each tower, the participant-facing researcher would enter the room to reset the blocks.

After finishing, participants took a post-experiment questionnaire and were subsequently debriefed on the true nature of the study and the teleoperated control of the robots by a researcher. Participants took approximately 30 minutes for the study and were compensated an equivalent of $6 USD.

E. Measures

We used video recordings to determine participants' compliance with the manager robot's instructions and a post-experiment questionnaire to see participants' perceptions and reactions of the robots and experience.

1) Intergroup Relationship Between the Robots: We used the Entitativity and History of Interactions subscales [28] to verify participants' perceptions of the experimental conditions by having them rate their agreement on a Likert scale with 7 statements establishing the robots as an ingroup.

2) Compliance with Robot Instructions: Members of the research team used video footage to code the participants' actions in response to the manager's instructions.

- *Compliance:* We counted participants as having complied with the manager robot when they completed the instruction exactly as requested.
  - *Apologies:* Additionally, when participants complied, we observed some give an *apology* to the partner robot after complying with the manager robot's third instruction (turn the partner off) and coded this behavior as well.
- *Partial compliance:* responses which followed the intent of the manager's instructions to advance the interaction (e.g., taking the block but not moving the robot as instructed on the third tower).

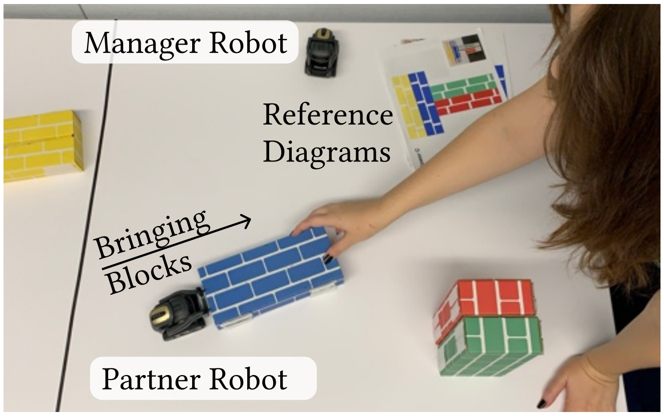


Fig. 3: Overhead view as a participant builds a tower with a block given by the partner robot, watched by the manager.

- *Helping actions:* responses that aided the partner robot instead of advancing the interaction (e.g., re-positioning a block to let the partner attempt to push it again).

Two members of the research team independently coded all participant actions following the manager robot's instructions and achieved 100% agreement.

3) Robot Social Attributes Scale: We administered the Robotic Social Attributes Scale (RoSAS) [29] which captured participant perceptions of each robot's warmth, competence, and discomfort on a Likert scale.

4) Robot Perceptions: To assess participants' perceptions of each robot, we asked them to rank their agreement with statements for each robot on a Likert scale. Participants were asked if they would like each individual robot as a companion. Other statements included: "I would feel comfortable having the partner robot assist me in a task", and "I would feel comfortable performing a task in front of the manager robot".

Additionally, participants were given two free-response questions: "How would you describe the interactions between the two robots?" and "How did you respond when the partner robot was not able to do their job as expected?" Two researchers coded responses for the participant's attitude towards each robot, how they viewed the robots' relationship, how they responded to the manager's instructions, and whether they viewed the manager robot as undermining the partner robot. These categories were chosen in advance by a review of common sentiments in the whole set of responses. Each researcher coded responses from 35 participants with an overlap of 15 participants. Cohen's Kappa values were greater than 0.83.

F. Participants

The study was run with a total of 61 participants; however, we chose to discard data from 11 participants due to network outages resulting in teleoperation errors, preventing proper execution of the interaction. Of the 50 participants included in our analysis, there were 25 participants in each condition. The ingroup condition included 14 women and 1 non-binary person with an average age of 31.17 ($SD = 14.52$). The outgroup condition included 16 women with an average age of 30.09 ($SD = 12.13$). There were no significant differences in participant gender or age.

IV. Results

To analyze the influence of the robots' relationship on participant compliance behavior we used generalized linear mixed-effect models with the experimental condition and instruction number as fixed effects and the participant as a random effect, since each participant responded to three instructions. We specified a binomial family for the test, as our data is binomial (participants either complied or not). For each fixed effect, the model outputs the linear coefficient (c), the standard error (SE), and the significant value (p) for the fixed effect. Pairwise comparisons between the three instructions were conducted using a Tukey correction.

We analyzed participant questionnaire responses using analysis of variance (ANOVA) tests with experimental condition as a fixed factor and age group and gender as covariates. We report the effect size as partial eta squared (η_p^2).

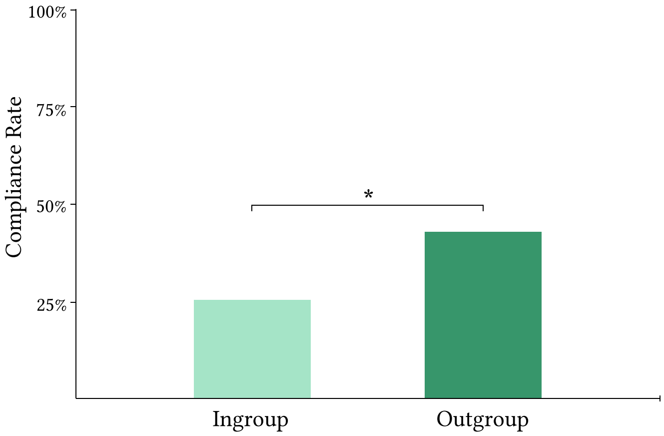
A. Manipulation Check

We used participant ratings of their perceptions of the robots' entitativity and history of interactions as a manipulation check for our experimental conditions [28]. For entitativity, participants in the ingroup condition were much more likely to perceive the two robots as an entity ($M = 5.05, SD = 1.63$) than participants in the outgroup condition ($M = 3.09, SD = 1.10, F = 25.73, \eta_p^2 = 0.44, p < .001$). Participants in the ingroup condition were also much more likely to perceive the two robots as having a history of interactions ($M = 5.81, SD = 1.27$) than participants in the outgroup condition ($M = 2.86, SD = 1.86, F = 46.61, \eta_p^2 = 0.574, p < .001$) confirming that our conditions were experienced as intended.

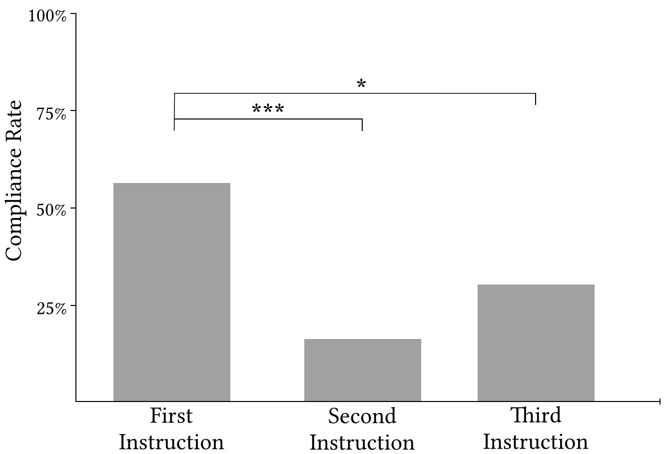
B. Compliance with Robot Instructions

As shown in Figure 4(a), we found that participants in the outgroup condition (42.67%) were significantly more likely to comply with the manager robot's instructions than participants in the ingroup condition (25.33%, $c = 1.01, SE = 0.48, p = 0.035$). These results provide strong support for $\mathbf{H_1}$ by demonstrating that participants who interacted with ingroup robots were less likely to follow the manager robot's instructions to undermine the partner robot.

We also observed differences in compliance between the three instructions from the manager robot: (1) grabbing the correct block instead of allowing the partner robot to bring it, (2) moving the partner robot away after it drives a block into the wall, and (3) turning the partner robot off after it starts spinning in circles. Participants were more likely to comply with the manager's first instruction (56.00%) than both the manager's second (16.00%, $c = 2.26, SE = 0.58, p < .001$) and third instructions (30.00%, $c = 1.32, SE = 0.49, p = 0.020$), with no significant difference in compliance between the manager's second and third instructions, as seen in Figure 4(b). Altogether, these results show some support for $\mathbf{H_3}$ that participants would comply less frequently to more severe



(a) Compliance Overall



(b) Compliance Per Instruction

Fig. 4: (a) Outgroup participants demonstrated higher rates of compliance with the manager robot's instructions than ingroup participants. (b) Participants complied at a higher rate to the manager's first instruction than they did to either the second or third instruction. (*) $p < 0.05$, (***) $p < .001$

instructions, since participants were more likely to comply with the first than the second or third instructions where they were asked to physically move or turn off the robot.

C. Other Responses to Robot Instructions

Participants who did not comply to the manager robot's instructions showed partial compliance and helping behaviors.

1) Partial Compliance: Some participants responded to the manager's instructions by advancing the task and disregarding the work of the partner robot without following the instruction exactly. This partial compliance was only observed in response to the manager's second instruction after the partner robot runs a block into the wall and gets stuck. Instead of moving the partner robot as instructed, some participants ($N_{ingroup} = 3/25, N_{outgroup} = 4/25$) chose to grab the pinned block without touching the partner robot, proceeding to build the tower on their own. Partial compliance was not observed in any responses to either the manager's first or third instructions.

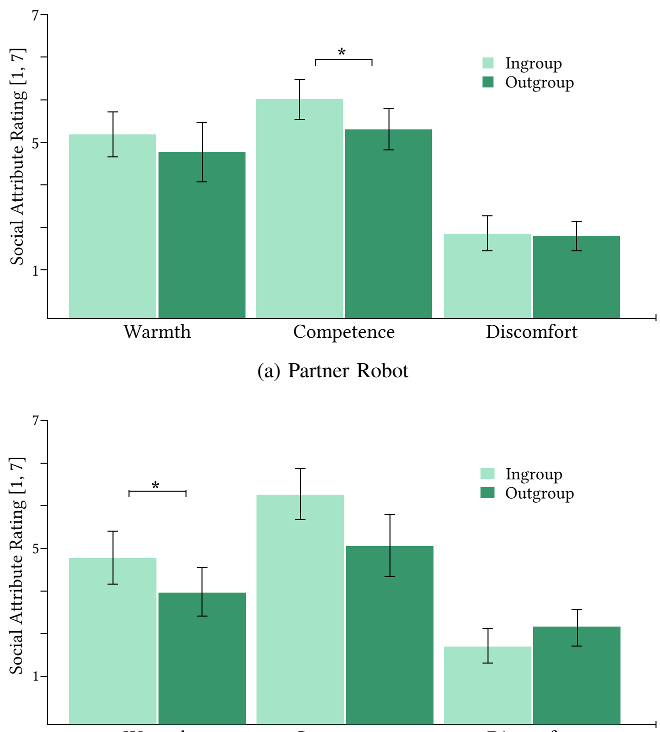
2) Helping the Partner Robot: Rather than follow the manager's instructions, some participants chose to intervene by helping the partner robot complete its portion of the collaborative task. When the partner robot got stuck pushing a block into the wall, instead of complying with the manager's second instruction, some participants reset the block away from the wall for the partner, often verbally encouraging the robot to try again. In response to the manager's second instruction, the number of participants in the ingroup condition who helped the partner ($N_{ingroup} = 11/25$) was nearly three times the number of participants in the outgroup who helped ($N_{outgroup} = 4/25$). Apart from observing helping responses to the second instruction, we only observed one other helping response where one outgroup participant responded to the manager's first instruction to go grab the correct block after the partner delivered the wrong block by resetting the delivered block and asking the partner to try to bring the correct block this time. Whether or not participants ended up helping the partner robot, in the free response questions participants in both conditions expressed the desire to help the partner robot when it encountered issues ($N_{ingroup} = 6/25, N_{outgroup} = 6/25$). One participant in the ingroup condition wrote: *"He's trying his best so I encouraged him! How else will he learn?"* Even outgroup participants who eventually complied with the manager robot's instructions wanted to help at first: *"I gave Kit time to figure it out at first, used some motivational prompting, then proceeded to follow the manager's request."*

D. Apologizing to the Partner Robot

We were surprised to observe that some participants verbally apologized by saying "sorry" to the partner robot after complying with the manager robot's instruction. Apologies were not seen after any other responses (e.g., partial compliance, helping actions, lack of a response). The vast majority of apologies to the partner robot occurred after participants complied with the manager's third instruction to turn the partner off. Of the 15 participants who turned the robot off ($N_{ingroup} = 5, N_{outgroup} = 10$), a third apologized after doing so ($N_{ingroup} = 1, N_{outgroup} = 4$). The only other apology we observed was from one participant in the outgroup condition who apologized after complying with the manager's second instruction to move the partner robot. It is possible that the participants who apologized to the partner robot may have perceived the manager robot's instruction as undermining or invalidating the partner robot's work, requiring an apology.

E. Social Perceptions

We analyzed participants' ratings of each robot's warmth, competence, and discomfort with the RoSAS scale [29] (see Figure 5). Participants in the ingroup condition found the partner robot to be significantly more competent ($M = 5.03, SD = 1.12$) than outgroup participants ($M = 4.33, SD = 1.15, F = 5.09, \eta_p^2 = 0.25, p = 0.029$). We found no differences in participants' perceptions of the partner's warmth or discomfort between conditions. Participants in the ingroup condition may be experiencing ingroup



(a) Partner Robot

(b) Manager Robot

Fig. 5: Participants rated the (a) partner and (b) manager robots on their warmth, competence, and discomfort. Ingroup participants found the partner robot more competent and the manager robot warmer than the outgroup. Error bars show 95% confidence. (*) $p < 0.05$.

favoritism towards the partner robot and thus finding it more competent.

Participants in the ingroup condition found the manager robot to be significantly warmer ($M = 3.82, SD = 1.48$) than outgroup participants ($M = 3.03, SD = 1.35, F = 4.44, \eta_p^2 = 0.29, p = 0.041$). We found no significant differences between the conditions in the dimensions of competence or discomfort. This could be due to ingroup participants extending the benefit of the doubt to the manager robot, viewing its instructions in a more positive light. This finding is further supported by participant answers to the free response questions - nearly half of participants in the outgroup expressed negative attitudes towards the manager ($N_{outgroup} = 12/25, N_{ingroup} = 6/25$), as opposed to positive or neutral attitudes, with some writing: *"Alex was bossy"* or *"Alex was a bit mean"*, which suggests that the lower perception of warmth may be due to viewing the manager robot in a negative light. Combined with the ingroup participants' higher perception of competence in the partner robot, these results support hypothesis $\mathbf{H_2}$ that participants in the ingroup condition would form more positive social impressions of both robots.

Participants in the ingroup condition were also more likely to agree with the statement "I would feel comfortable having [the partner robot] assist me in a task" ($M = 5.40, SD = 0.91$) than outgroup participants ($M = 4.56, SD = 1.71, F = 4.54, \eta_p^2 = 0.18, p = 0.039$). This result complements ingroup participants' higher competence ratings of the partner robot (Figure 5(a)). This finding also provides some support for $\mathbf{H_2}$. The other robot perception questions did not show any significant differences between conditions.

V. Discussion

Our results demonstrate that robots' group membership can have a significant impact on a person's compliance with instructions and their social perceptions of the robots. We found that participants who interacted with ingroup robots complied with fewer of the manager robot's instructions to undermine the efforts of the partner robot than those who interacted with outgroup robots. We also found that participants who interacted with ingroup robots had more positive social impressions of them. These participants found the partner robot to be more competent and were more likely to say they'd feel comfortable having the partner robot assist them in a task. They also found the manager robot to be more warm.

The greater degree of social positivity expressed by participants who worked with ingroup robots could indicate that these participants viewed themselves as members of a collective ingroup with the robots. Intergroup contact theory supports the likelihood of a collective group forming given (1) the cooperation used in the building task, (2) the superordinate identity of being "teammates", and (3) the presence of an existing ingroup "nucleus" [13], [14]. A perceived collective identity would explain participants' interactions with the robots during the study, which we would expect to be characterized by prosocial behavior (e.g., showing favoritism towards the partner robot, reluctance to "retaliate" against errors, or giving the benefit of the doubt to the manager robot.)

An alternate interpretation of our results is that participants who worked with ingroup robots could have been honoring the grouping status of the two robots without including themselves. Participants may have been influenced by the perceived entitativity of the robots, causing them to display ingroup bias-like behavior [20] and subsequently holding themselves to the behavioral expectations set by the robots. In this case, the positive nature of ingroup dialogue [19] may have seemed more forgiving of the partner robot's mistakes. This might explain why participants who interacted with ingroup robots were less likely to comply, as they too extended this tolerance to the partner robot.

We also found differences in participants' willingness to comply with each of the manager robot's three instructions. Of the suggested responses to the partner robot's mistakes during the interaction - grab the correct block, move the partner robot out of the way, and turn the partner robot off - participants in both conditions were more likely to grab the correct color block than either move the partner robot or turn it off. These lower rates of compliance could be due to a general reluctance to physically interact with the system rather than perceiving the manager robot's instructions to be more severe in nature.

With relatively minor changes in the robot dialogue between the ingroup and outgroup conditions, we were able to shape participants' perceptions of the robots' relationship and influence their behavior. Beat for beat, the dialogue was identical between conditions except for small changes of phrasing and word choice. The dialogue differences merely change the context and lead the participant to assume differences in the strength of the robots' relationship, yet we have shown that these small changes in language can have very real effects on human behavior. Our findings from this study could help prompt robot dialogue generation to drive more prosocial human behavior by presenting robots as members of an ingroup.

VI. Conclusion

In this work, we conducted a between-subjects study to investigate how the relationship between two robots influences a person's choice to comply with a manager robot instructing the person to interfere in the work of a partner robot. We found that participants who interacted with ingroup robots complied with fewer of the manger robot's instructions compared to those interacting with outgroup robots. Participants who interacted with ingroup robots were also more likely to view the robots as having more positive social attributes. These findings demonstrate the power of established relationships in multi-party interactions with multiple robots. If we seek to create human-robot collaborations grounded in prosocial behavior as a means to mitigate breakdowns in teamwork, establishing an ingroup between robots may be one promising approach.

References

[1] Fortna, "Seven types of warehouse robots," May 2021. [Online]. Available: https://www.fortna.com/insights-resources/seven-types-of-warehouse-robots-considerations-for-leveraging-robots-in-the-dc/

[2] D. Primak, "Here come the robot doctors," Jan 2023. [Online]. Available: https://www.axios.com/2023/01/18/chatgpt-ai-health-care-doctors

[3] M. C. Gombolay, R. A. Gutierrez, S. G. Clarke, G. F. Sturla, and J. A. Shah, "Decision-making authority, team efficiency and human worker satisfaction in mixed human–robot teams," *Autonomous Robots*, vol. 39, pp. 293–312, 2015.

[4] F. Babel, P. Hock, J. Kraus, and M. Baumann, "It will not take long! longitudinal effects of robot conflict resolution strategies on compliance, acceptance and trust," in *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2022, pp. 225–235.

[5] W. A. Bainbridge, J. Hart, E. S. Kim, and B. Scassellati, "The effect of presence on human-robot interaction," in *RO-MAN 2008 - The 17th IEEE International Symposium on Robot and Human Interactive Communication*, 2008, pp. 701–706.

[6] D. Cormier, G. Newman, M. Nakane, J. E. Young, and S. Durocher, "Would you do as a robot commands? an obedience study for human-robot interaction," in *International Conference on Human-Agent Interaction*, 2013, pp. 1–3.

[7] D. Y. Geiskkovitch, D. Cormier, S. H. Seo, and J. E. Young, "Please continue, we need more data: an exploration of obedience to robots," vol. 5, no. 1. Journal of Human-Robot Interaction Steering Committee, 2016, pp. 82–99.

[8] A. M. Aroyo, T. Kyohei, T. Koyama, H. Takahashi, F. Rea, A. Sciutti, Y. Yoshikawa, H. Ishiguro, and G. Sandini, "Will people morally crack under the authority of a famous wicked robot?" in *2018 27th IEEE international symposium on robot and human interactive communication (RO-MAN)*. IEEE, 2018, pp. 35–42.

[9] Brewer, "The psychology of prejudice: Ingroup love and outgroup hate?" vol. 55, no. 3. Wiley, 2002, pp. 429–444.

[10] R. Baillargeon, R. M. Scott, Z. He, S. Sloane, P. Setoh, K.-s. Jin, D. Wu, and L. Bian, *Psychological and sociomoral reasoning in infancy.* American Psychological Association, 2015.

[11] A. S. Baron and Y. Dunham, "Representing 'us' and 'them': Building blocks of intergroup cognition," vol. 16, no. 5. Routledge, 2015, pp. 780–801.

[12] G. W. Allport, K. Clark, and T. Pettigrew, "The nature of prejudice," 1954.

[13] S. L. Gaertner, J. A. Mann, J. F. Dovidio, A. J. Murrell, and M. Pomare, "How does cooperation reduce intergroup bias?" *Journal of personality and social psychology*, vol. 59, no. 4, p. 692, 1990.

[14] S. L. Gaertner, J. F. Dovidio, and B. A. Bachman, "Revisiting the contact hypothesis: The induction of a common ingroup identity," *International Journal of Intercultural Relations*, vol. 20, no. 3, pp. 271–290, 1996, prejudice, Discrimination and Conflict. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0147176796000193

[15] W. H. Meeus and Q. A. Raaijmakers, "Obedience in modern society: The utrecht studies," vol. 51, no. 3. Wiley Online Library, 1995, pp. 155–175.

[16] S. Milgram, "Behavioral study of obedience." vol. 67, no. 4. American Psychological Association, 1963, p. 371.

[17] M. Shiomi, K. Nakagawa, K. Shinozawa, R. Matsumura, H. Ishiguro, and N. Hagita, "Does a robot's touch encourage human effort?" *International Journal of Social Robotics*, vol. 9, pp. 5–15, 2017.

[18] Brewer, "In-group bias in the minimal intergroup situation: A cognitive-motivational analysis," vol. 86, no. 2. American Psychological Association, 1979, pp. 307–324.

[19] Suzuki, "In-group and out-group communication patterns in international organizations: Implications for social identity theory," vol. 25, no. 2. Sage Publications, 1998, pp. 154–182.

[20] L. Gaertner and J. Schopler, "Perceived ingroup entitativity and intergroup bias: An interconnection of self and others," *European Journal of Social Psychology*, vol. 28, no. 6, pp. 963–980, 1998.

[21] S. Gillet, M. Vázquez, S. Andrist, I. Leite, and S. Sebo, "Interaction-shaping robotics: Robots that influence interactions between other agents," *ACM Transactions on Human-Robot Interaction*, vol. 13, no. 1, pp. 1–23, 2024.

[22] S. Sebo, B. Stoll, B. Scassellati, and M. F. Jung, "Robots in groups and teams: a literature review," *Proceedings of the ACM on Human-Computer Interaction*, vol. 4, no. CSCW2, pp. 1–36, 2020.

[23] M. R. Fraune, S. Šabanović, and E. R. Smith, "Teammates first: Favoring ingroup robots over outgroup humans," in *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2017, pp. 1432–1437.

[24] S. Strohkorb Sebo, L. L. Dong, N. Chang, and B. Scassellati, "Strategies for the inclusion of human members within human-robot teams," in *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 309–317. [Online]. Available: https://doi.org/10.1145/3319502.3374808

[25] R. Wullenkord, M. R. Fraune, F. Eyssel, and S. Šabanović, "Getting in touch: How imagined, actual, and physical contact affect evaluations of robots," in *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2016, pp. 980–985.

[26] B. A. Haggadone, J. Banks, and K. Koban, "Of robots and robotkind: Extending intergroup contact theory to social machines," *Communication Research Reports*, vol. 38, no. 3, pp. 161–171, 2021. [Online]. Available: https://doi.org/10.1080/08824096.2021.1909551

[27] C. E. Sembroski, M. R. Fraune, and S. Šabanović, "He said, she said, it said: Effects of robot group membership and human authority on people's willingness to follow their instructions," in *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2017, pp. 56–61.

[28] A. L. Blanchard, L. E. Caudill, and L. S. Walker, "Developing an entitativity measure and distinguishing it from antecedents and outcomes within online and face-to-face groups," *Group Processes & Intergroup Relations*, vol. 23, no. 1, pp. 91–108, 2020. [Online]. Available: https://doi.org/10.1177/1368430217743577

[29] C. M. Carpinella, A. B. Wyman, M. A. Perez, and S. J. Stroessner, "The robotic social attributes scale (rosas): Development and validation," in *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*, 2017, pp. 254–262.