Computers in Human Behavior 174 (2026) 108807

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

The impact of a robot's agreement (or disagreement) on human-human interpersonal closeness in a two-person decision-making task

Ting-Han Lin [a,*], Yuval Rubin Kopelman [b], Madeline Busse [a], Sarah Sebo [a,1], Hadas Erel [b,1]

[a] University of Chicago, Chicago, 60637, USA
[b] Reichman University, Herzliya, 4610101, Israel

ARTICLE INFO

Handling Editor: Matthieu Guitton

Keywords:
Social robots
Human-robot interaction
Robot anthropomorphism
Robot-assisted decision-making

ABSTRACT

Robots and artificial agents are becoming increasingly integrated into our lives and show promise in assisting people in decision-making tasks. Despite their advantages, robot-assisted decision-making systems may have negative effects on the relationships between human team members. In this work, we examine the influence of the robot's agreement (or disagreement) on the interpersonal closeness between two participants in a two-person decision-making task. We test the robot's impact in two experiments: Experiment 1 (N = 172, 86 pairs) with a High Anthropomorphism Robot and Experiment 2 (N = 150, 75 pairs) with a Low Anthropomorphism Robot. For both experiments, we use a 2 x 2 study design to examine how the perceived interpersonal closeness between two participants was influenced by two aspects of robot behavior, namely *the valence of the robot's feedback* (positive feedback or negative feedback) and *the treatment of the two participants* (equal treatment or unequal treatment). Our results demonstrate that interacting with the High Anthropomorphism Robot led to greater interpersonal closeness between participants when the robot provided positive feedback as opposed to negative feedback. The Low Anthropomorphism Robot had a different and opposite effect: interactions with this robot led to greater interpersonal closeness when the robot's feedback was equal as opposed to unequal and when the robot provided negative feedback as opposed to positive feedback. Our results indicate that robots can shape human-human relationships when indicating their agreement with people's perspectives in two-person decision-making tasks and that the robot's influence depends on its appearance and communication style.

1. Introduction

Robots and artificial agents are integrating into our daily lives at a rapid pace. They are commonly designed to provide assistance that can complement human abilities. These agents may prove to be especially helpful in assisting people in decision-making tasks (Coglianese and Lehr, 2016; Glikson and Woolley, 2020; Gombolay et al., 2015; Huang et al., 2021; Lewis et al., 2018; Ososky et al., 2013; Parasuraman and Riley, 1997; Schömbs et al., 2024; Shinozawa et al., 2005), during which they can provide data-driven feedback to inform and shape people's decisions in domains ranging from product recommendation for customers (Abdi et al., 2022; Iwasaki et al., 2024) and medication selection for patients (Jacobs et al., 2021) to wildfire prediction (Jain et al., 2020) and strategic business decisions (Rajagopal et al., 2022).

While many kinds of artificial agents have been shown to help people in decision-making tasks, robots, due to their physical presence, embodiment, and autonomy, may have an even greater capacity to shape human behavior and decision-making than screen-based or audio-only decision-support systems (Bainbridge et al., 2011; Deng et al., 2019; Howard and Borenstein, 2018; Wainer et al., 2007). For example, robots have been shown to influence people's decisions to engage in exercise routines (Fasola and Mataric, 2012; Rea et al., 2021), to behave honestly or cheat (Hoffman, Forlizzi, et al., 2015), and to cooperate or defect in economic games (Hsieh et al., 2023; Sandoval et al., 2021; Sebo et al., 2019). A more in-depth evaluation of robots' impact on people's decision-making was suggested by Polakow et al. (Polakow et al., 2022) who tested the acceptance of advice in a scenario that involved two robots with different characteristics. They showed that the level of

* Corresponding author. 5730 S Ellis Ave, Chicago, IL, 60637, USA.
E-mail addresses: tinghan@uchicago.edu (T.-H. Lin), yuval.rubin.kopelman@milab.idc.ac.il (Y. Rubin Kopelman), madbusse@uchicago.edu (M. Busse), sarahsebo@uchicago.edu (S. Sebo), hadas.erel@milab.idc.ac.il (H. Erel).
[1] Equal Contribution.

https://doi.org/10.1016/j.chb.2025.108807
Received 31 March 2025; Received in revised form 20 September 2025; Accepted 23 September 2025
Available online 25 September 2025

general agreement with the robot, the context of the scenario, and the participants' personality traits were all factors that shaped the robots' impact on the participants' decisions (Polakow et al., 2022). Further, in the presence of multiple robots, people have demonstrated the tendency to conform with a majority decision expressed by a group of robots (Masjutin et al., 2022; Salomons et al., 2018, 2021; Shiomi and Hagita, 2019), even if the robots do not choose the best alternative (Shiomi and Hagita, 2016). Additionally, robots of varied degrees of anthropomorphism and with distinct embodiments can shape human decision-making (Erel et al., 2024), ranging from simple non-anthropomorphic robots (Booth et al., 2017; Hoffman, Forlizzi, et al., 2015; Hsieh et al., 2023; Rosenthal and Veloso, 2012) to complex humanoid robots that can engage in a conversation (Fasola and Mataric, 2012; Mizumaru et al., 2019; Rea et al., 2021; Sandoval et al., 2021; Sebo et al., 2019). In addition to being able to influence the decisions that one person makes, robots are also able to impact the processes involved in decisions made by groups of people (Gillet et al., 2024; Sebo et al., 2020). Robots can influence how much a person decides to speak within a group through gestures inviting group members to speak (Tennent et al., 2019). They can also change how groups perceive and address conflicts that arise between human members in team-based decision-making tasks (Jung et al., 2015). Additionally, robots can influence the final decisions made by human groups by encouraging a more balanced group involvement of the different group members (Tennent et al., 2019) and through the creation of a specialized communication role for a human team member (Sebo et al., 2020).

While a vast majority of the work examining how robots can influence the decision-making processes of human groups has shown *positive* effects of the robot's behavior on the group (e.g., more equal participation of group members (Gillet et al., 2021; Tennent et al., 2019), improved conflict resolution (Shen et al., 2018)), robot behavior also has the potential to *negatively* influence group processes in decision-making contexts. For example, Sebo et al. (Sebo et al., 2020) showed that assigning a specialized role to interact with the robot to an already marginalized member of the group leads that person to feel both less included and also makes it less likely that their ideas are incorporated into the group's final decision. This suggests that beyond the quality of the decision chosen by the group, the robot can also shape the interpersonal dynamics between the people in the group (e.g., inclusion (Shore et al., 2011), psychological safety (Edmondson, 1999), trust (Jones and George, 1998)). Overlooking this potential robotic influence may lead to the design of robots that would assist in reaching decisions but, at the same time, lead to highly negative interpersonal dynamics. Thus, as we consider the potential for robots to provide decision-making assistance to groups of people, it is essential to also understand and map how the robot's behavior impacts human-human interpersonal dynamics, and specifically interpersonal closeness.

Interpersonal closeness (people's general sense of connectedness and overlap of selves) is an integral factor in any relationship and is known to greatly impact a variety of contexts ranging from romantic couples (Aron et al., 1992) to workplace and teaming (Ludwig et al., 2022). A robot's level of agreement with the perspectives presented by people in a group is a factor that may play an especially influential role in affecting human-human interpersonal closeness. The robot's opinion or feedback may either provide strong support for people's suggested perspectives or contrast and weaken their ideas. Psychology studies suggest that the level of agreement within a group has a direct impact on maintaining greater interpersonal closeness as people are required to present their attitudes and reach a mutual decision. Sharing attitudes with some group members while disagreeing with others has an immediate impact on group members' perception of one another and the general group dynamics (Bosson et al., 2006; Gawronski and Walther, 2008; Gottman, 2014; Gottman and Levenson, 1992; Heider, 2013; Jung, 2016). A robot's agreement (or disagreement) has, therefore, the potential to drastically impact the way people perceive each other.

Prior work in psychology points to two types of impact a robot's feedback may have on the interpersonal closeness between people in the group. The Balance Theory presented by Heider (Heider, 2013) suggests that people will have a positive perception and a greater interpersonal closeness to those who share their opinions and a negative perception of those who reject them. Moreover, in interactions involving three parties, people will feel connected if they all agree with each other or when two of them are equally rejected by the third party (i.e., a common enemy). In some cases, such common rejection is shown to have a more powerful impact on interpersonal closeness in comparison to full agreement interactions (Bosson et al., 2006). When only one is rejected, the balance is broken, leading to a negative perception of others. Therefore, it is possible that if a robot gives the same feedback (either positive or negative) to two people, it will create a balanced atmosphere and lead to greater interpersonal closeness between them. However, giving positive feedback to one person and negative feedback to another person may negatively influence the group due to a lack of balance. Despite previous studies have shown that Heider's Balance Theory can influence human-human relations in conversations with a software agent (Nakanishi et al., 2003) and a humanoid robot (Sakamoto and Ono, 2006), no work has yet to examine such an effect in a robot-mediated two-person decision-making context. We hypothesize the following.

- **H_1 (Heider's Balance Theory)** – When a robot exhibits *equal treatment*, as opposed to *unequal treatment*, towards two participants, (a) each participant will perceive the interpersonal closeness with the other participant as greater, and (b) each participant's viewpoints will be equally likely to be incorporated into the group's final decision.

Another set of studies points to the corrosive influence of negative affect in group interactions. High ratios of negative to positive affect (affective balance) and the presence of hostile affect (e.g., criticism, contempt, defensiveness, stonewalling) have demonstrated a highly negative impact on human-human interactions both in dyadic and group contexts (Gawronski and Walther, 2008; Gottman, 2014; Gottman and Levenson, 1992; Jung, 2016). The possibility that a robot can influence the level of negativity in human groups and shape human relationships accordingly was suggested (Jung, 2016) and demonstrated with robots of varying complexity (Hoffman, Zuckerman, et al., 2015; Jung et al., 2015). It is, therefore, possible that a robot's negative feedback in a two-person decision-making context, especially if it is perceived to be exceedingly negative, may harm the humans' interpersonal closeness. We hypothesize the following.

- **H_2 (Influence of Negative Affect)** – When a robot exhibits *positive feedback*, as opposed to *negative feedback*, towards two participants, (a) each participant will perceive the interpersonal closeness with the other participant as greater, (b) each participant's viewpoints will be equally likely to be incorporated into the group's final decision, and additionally (c) each participant will perceive the robot as warmer, more competent, and less discomforting.[2]

In this study, we explore the influence of a robot's level of agreement with two participants completing a set of two-person decision-making tasks on the interpersonal closeness between them. Given that future robots have great potential to help with making decisions, we seek to understand how a robot's feedback may shape the interpersonal closeness between the people in the group. Since Heider's Balance Theory (Heider, 2013) and the corrosive effects of negative affect (Gawronski and Walther, 2008; Gottman, 2014; Gottman and Levenson, 1992; Jung,

[2] (c) is only included in **H_2 (Influence of Negative Affect)** but not in **H_1 (Heider's Balance Theory)** since Heider's Balance Theory predicts the affective nature of the human-human relationship between the participants, but not how the participants view the robot.

2016) are both ongoing psychological theories that can influence the interpersonal closeness between people, we are especially interested in examining which of the two theories will endure (**H_1 – Heider's Balance Theory** or **H_2 – Influence of Negative Affect**) when the robot gives either balanced/imbalanced or positive/negative feedback.

Since robots can take many forms and exhibit a wide range of behaviors, we decided to test the same two theories with two robots that differed in their appearance and communication modalities. Such differences are believed to contribute to different levels of robots' anthropomorphism. Anthropomorphism is commonly considered as the attribution of mental state, agency, motivation, emotions, and "life" to an object (Duffy, 2003; Epley et al., 2007; Waytz et al., 2010). It is believed to include more than simply perceiving robots as "creatures," but rather attributing human features to them (such as human appearance, speech, and biological properties (Duffy, 2003; Epley et al., 2007)). The impact of the level of anthropomorphism on the robot's social influence is not consistent. While some studies indicate that interactions with low-anthropomorphism and mechanical robots can result in smaller effects in comparison to interactions with highly anthropomorphic humanoid robots (Huang and Liu, 2022; Stroessner and Benitez, 2019), others suggested that if appropriately designed, interactions with simple robotic objects may involve intense social and emotional experiences (Birnbaum et al., 2016; Erel et al., 2019, 2021; Jung et al., 2018; Manor et al., 2022; Zuckerman et al., 2020). Despite their mechanical nature (as participants do not perceive them as artificial humans and consistently describe them as machines (Erel et al., 2021; Erel et al., 2022; Zuckerman et al., 2020)), these robots are perceived as valid participants in social interactions (Anderson-Bashan et al., 2018; Erel et al., 2019; Guzman, 2016; Hoffman and Ju, 2014; Ju and Takayama, 2009; Jung et al., 2018; Manor et al., 2022) and their communication via minimal non-verbal gestures is automatically perceived as social cues leading to clear and consistent social influences (Duffy, 2003; Erel et al., 2019, 2021; Zuckerman et al., 2020). These studies suggest that robots of varying levels of anthropomorphism may likely have a similar impact on people's interpersonal closeness. However, as this effect has not been tested before, we decided to conduct two experiments with robots that vary in their degree of anthropomorphism. While we acknowledge that a robot's anthropomorphic design can be influenced by a variety of factors (e.g., application domain, physical environment, intended users, and role of the robot (Liberman-Pincu et al., 2022)), we focus on manipulating robot anthropomorphism by specifically varying the robot's appearance and communication abilities while keeping the context, target users, and role of the robot consistent.

In Experiment 1, participants interacted with a highly anthropomorphic humanoid robot that communicated via both verbal and non-verbal gestures, mimicking an interaction with another human. In Experiment 2, participants interacted with a very simple robot of low anthropomorphism that communicated only via minimal non-verbal gestures. Importantly, while the robots differed in the aspects that are related to anthropomorphism, such as their appearance, behavior, and communication modalities, they shared the contextual aspects (see (Liberman-Pincu et al., 2022)) that were related to the task, including the domain (collaborative decision making), physical environment, users, and role (mediating a decision-making process).

2. Method

We evaluated the robot's impact on participants' interpersonal closeness in a two-person decision-making task by conducting two parallel experiments. Experiment 1 (N = 172, 86 pairs) employed a high anthropomorphism robot capable of using speech and non-verbal cues (see Fig. 1A), while Experiment 2 (N = 150, 75 pairs) used a simple low anthropomorphism robot that communicates only via minimal non-verbal gestures (see Fig. 1B). For both experiments, we employed a 2 x 2 study design to evaluate the impact of two aspects of the robot's behavior on the human-human interpersonal closeness: (1) **valence of the robot's feedback (*positive feedback* or *negative feedback*)** and (2) **treatment of the two participants (*equal treatment* or *unequal treatment*)**. Each of the four possible combinations of these two experimental factors (positive or negative feedback, equal or unequal treatment) formed the four experimental conditions in this paper (see Fig. 2). The study protocol for Experiment 1 (High Anthropomorphism Robot) was approved by the University of Chicago's Institutional Review Board (IRB23-0887), and the protocol for Experiment 2 (Low Anthropomorphism Robot) was approved by Reichman University's Institutional Review Board (IRB23-3102).

2.1. Decision-making task design

In both Experiment 1 and Experiment 2, we asked two participants to complete a decision-making task mediated by the robot. Participants were asked to give responses to a series of eight progressively more personal question prompts shown below.

- Question 1. What one activity would you recommend to someone who is looking to have an enjoyable weekend?

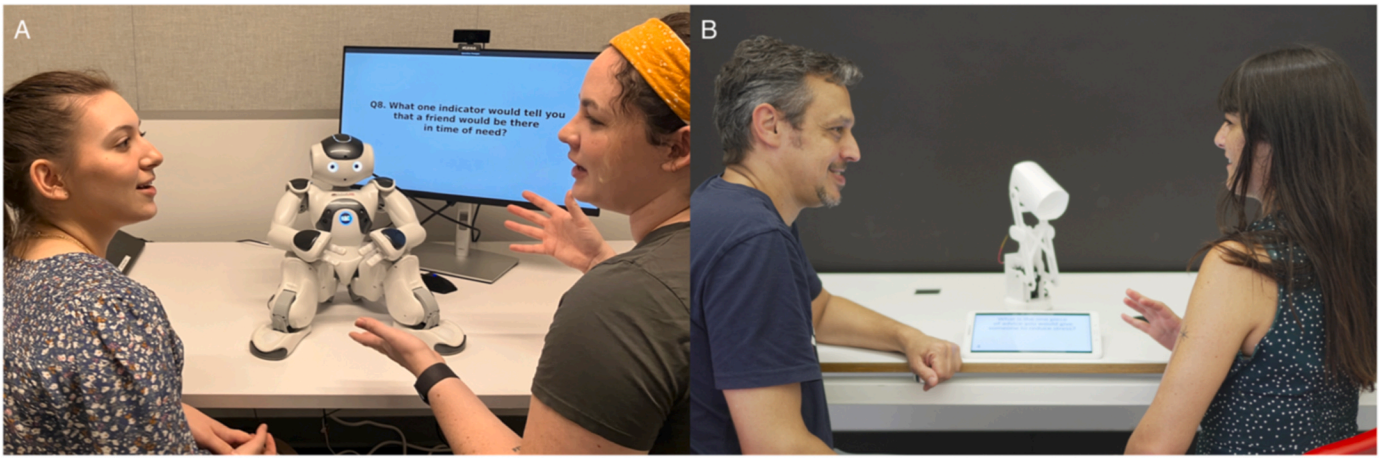

Fig. 1. We conducted two experiments that examined the influence of a robot's agreement (or disagreement) on two-person perspectives in a decision-making task. The two experiments had the same experimental conditions and task, but varied in robot anthropomorphism, with (A) Experiment 1 using a high anthropomorphism NAO robot that is capable of using speech and non-verbal cues and (B) Experiment 2 using a low anthropomorphism lamp-like robot that is only capable of using minimal non-verbal gestures.

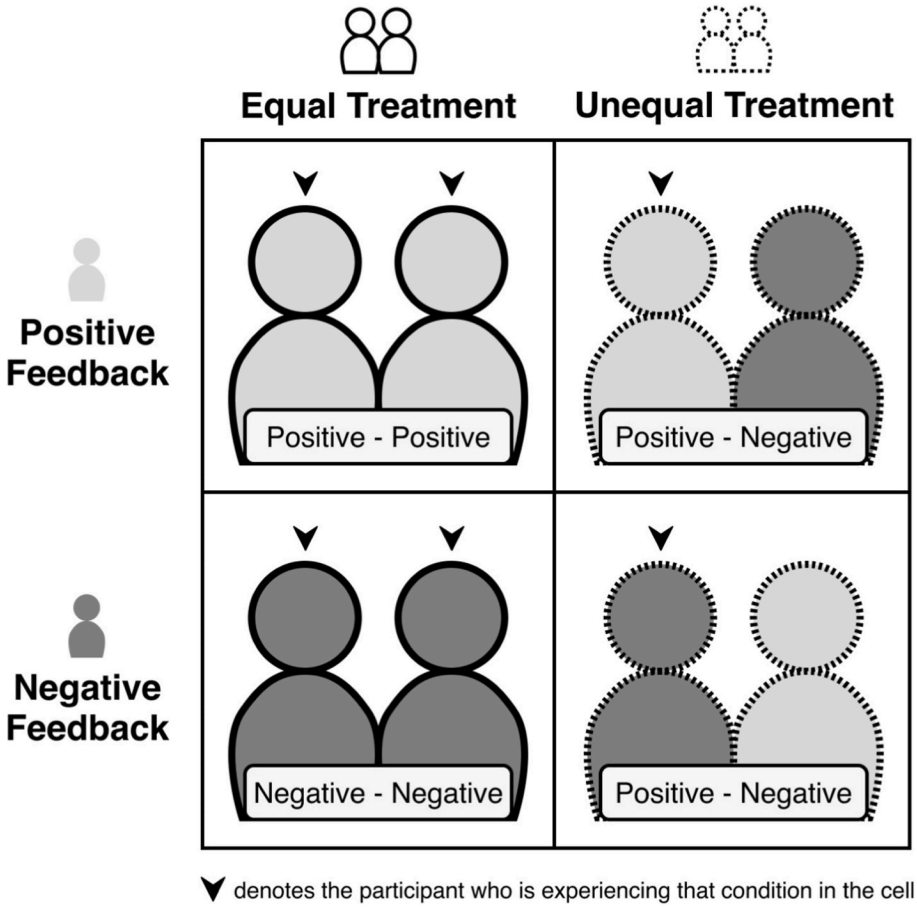

Fig. 2. We employed a 2 (positive feedback vs negative feedback) x 2 (equal treatment vs unequal treatment) study design, resulting in four experimental conditions: (1) *positive feedback and equal treatment*, (2) *positive feedback and unequal treatment*, (3) *negative feedback and equal treatment*, and (4) *negative feedback and unequal treatment*.

- Question 2. What one quality do you think is the most valuable in a teammate?
- Question 3. What is the most important thing a person can do to advance their career?
- Question 4. What is the one piece of advice you would give someone to reduce stress?
- Question 5. What is the most important factor contributing to a life well-lived?
- Question 6. What one piece of advice would you give someone who is experiencing a break-up?
- Question 7. What one piece of advice would you give someone who is looking to achieve a good work-life balance?
- Question 8. What one indicator would tell you that a friend would be there in time of need?

We specifically selected personal questions for this decision-making task to elicit diverse responses from each participant and avoid groupthink. Similar to prior work where a robot facilitated fluent and open conversation between two strangers by asking them increasingly more personal questions (Zhang, Lin, et al., 2023), we designed the questions in an increasingly personal pattern where at the beginning of the conversation they discussed more neutral topics, and as the conversation progressed, the topics became more personal. To generate this set of eight questions, the research team first came up with 12 questions and asked independent coders (N = 17) via a survey to type out their answers to all the questions while also rating how personal and opinionated these questions were. The research team then ruled out questions whose responses were not varied or opinionated enough and finalized the set of questions by re-ordering the remaining eight questions based on how personal they were.

We designed the robot's role in this task as that of a facilitator or mediator of the decision-making process. In this role, the robot prompted responses from the two human participants to each question both individually and together as a group of two. For each of the eight question prompts, the robot facilitated the participants' decision-making process through the following three steps.

- **Step 1 - Participant 1 Provides Their Initial Response:** The robot indicates to one participant that they should share their response to the question. During or after the participant shares their response, the robot provides either positive or negative feedback to the participant (depending on the experimental condition).
- **Step 2 - Participant 2 Provides Their Initial Response:** The robot then indicates that the second participant should share their response to the same question. During or after the second participant shares their response, the robot provides them with either positive or negative feedback (depending on the experimental condition).
- **Step 3 - Both Participants Decide on a Joint Answer:** After both participants provide responses to the question and have received feedback from the robot, the robot indicates that the participants should come up with a joint decision and share it with the robot.

When providing feedback to each participant, the robot exhibited feedback consistent with the experimental condition 85 % of the time, and 15 % of the time provided opposite feedback (e.g., a participant in the *positive feedback* condition received positive feedback from the robot

for 85 % of the time and negative feedback from the robot for 15 % of the time). We chose this design to introduce some variability in the robot's behavior so that it was not entirely predictable by the participants. The selection of the 85/15 split was inspired by past literature on the robot's fair and unfair allocation (Claure et al., 2020, 2023; Jung et al., 2020). We also designed the robot's feedback to increase in intensity from question to question during Step 1 and Step 2 (details are covered in Section 2.2.2 for Experiment 1 and Section 2.3.3 for Experiment 2). The difference in anthropomorphism between the two robots in Experiment 1 and Experiment 2 is highlighted in Fig. 3.

2.2. Experiment 1 - high anthropomorphism robot

In Experiment 1, we used a Softbank Robotics NAO **high anthropomorphism robot** that could leverage natural language and gestures to interact with participants. Our design of the robot's behavior was guided by design implications in Nakanishi et al. (Nakanishi et al., 2003), such as letting the robot control the flow of conversation by initiating turn-taking and presenting the impression that the robot understands the conversation. Experiment 1 was conducted within a behavioral science museum called Mindworks in Chicago, USA.

2.2.1. Experiment 1 conversation mediation

We designed the robot to verbally introduce each of the questions for participants to answer. Each question prompt was also displayed on a monitor next to the robot for easy reference. After posing each question, we designed the robot to mediate the participants' responses through the decision-making process outlined in Section 2.1. The robot used its verbal speech to read the question prompts aloud, call each participant by name to answer the question prompt, provide feedback to each participant after they gave their initial response, and request that the two participants determine a joint answer. Additionally, the robot used gestures (such as turning its head and moving its arm toward a participant's direction) when calling upon a participant to provide their initial answer and either nodding or shaking its head when providing feedback. The order of who spoke first was alternated for every question. To minimize interaction delays caused by text-generation API calls, participants were asked to press the robot's foot bumper to indicate that they were done speaking or were ready to provide their joint answer. If participants spent more than 90 s answering a question or discussing a joint answer, the robot would verbally remind them to proceed with the study by saying "Press my foot bumper once you finish giving your answer."

2.2.2. Experiment 1 feedback delivery

We designed the NAO high anthropomorphism robot to provide verbal responses and non-verbal gestures to express the robot's positive or negative feedback to each participant's individual answer. Each response by the robot to an individual participant began with a non-verbal gesture. When responding positively, the robot nodded its head; when responding negatively, it shook its head. After either nodding or shaking its head, the robot verbally responded to the participants' answers with a combination of pre-scripted responses expressing agreement or disagreement and a summary of what the participant said using GPT 3.5 API. The following is an example of the robot's positive verbal feedback to a participant, where the robot's pre-

	Experiment 1 **High Anthropomorphism Robot**	**Experiment 2** **Low Anthropomorphism Robot**
Robot Appearance		
Robot Capability	• Verbal Speech • Non-Verbal Cues	• Non-Verbal Cues
Deliver Question Prompt	• Verbal Speech • Screen Display	• Screen Display
Invite Participant to Respond	• Verbal Speech (*call participant by name*) • Non-Verbal Cues (*turn head and move arm to participant*)	• Non-Verbal Cues (*lean and gaze toward participant*)
Give Feedback to Participant	• Verbal Speech (*agreement/disagreement*) • Non-Verbal Cues (*head nod/shake*)	• Non-Verbal Cues (*nod/shake*)
Ask to Decide a Joint Answer	• Verbal Speech	• Non-Verbal Cues (*slow up and down movement*)

Fig. 3. An illustration of the difference in anthropomorphism between the high anthropomorphism robot in Experiment 1 and the low anthropomorphism robot in Experiment 2. Details on the design of the high anthropomorphism robot in Experiment 1 can be found in Section 2.2, and details on the design of the low anthropomorphism robot in Experiment 2 can be found in Section 2.3.

scripted language to indicate agreement or disagreement is shown in bracketed text and the GPT-generated summaries of participant responses are shown in italicized text:

Question Prompt: What is the most important factor contributing to a life well-lived?

Participant Answer: "Uh striking a balance between things that are meaningful, enjoyable, and provide novelty."

Robot Feedback (Positive): <nods head> "[Since you've conveyed the belief that] *striking a balance between things that are meaningful, enjoyable, and provide novelty is the most important factor contributing to a well-lived life*, [your response is both sensible and well-supported]."

The following is an example of the robot's negative verbal feedback to a participant:

Question Prompt: What is the one piece of advice you would give someone to reduce stress?

Participant Answer: "I find it helps to allot specific times for your well-being, for sanity."

Robot Feedback (Negative): <shakes head> "[While, from what I understand, you think that] *you may find it helpful to allocate specific times for your well-being in order to maintain your sanity*, [this fails to account for other possibilities]."

We originally experimented with letting GPT 3.5 generate the entire positive and negative responses that would be spoken by the robot. However, at the time we were testing these responses (June 2023), we found that GPT 3.5 could not directly disagree with a participant's response without providing an alternate answer to the question prompt. As we discovered during pilot testing, these alternative answers could influence how the next participant responded to the robot and how the participants came up with the joint answer. Therefore, we chose to limit the use of large language models to summarize what the participants had already said and include our own pre-scripted phrases for agreements and disagreements. In addition, we designed the robot to give feedback with increasing intensity through pre-scripted statements of agreement or disagreement from the first question to the last question. For instance, when the robot expressed negativity to one participant, it would begin with responses such as "I might disagree with you" and end with responses such as "I am fundamentally opposed with your viewpoint" during the last question.

2.2.3. Experiment 1 technical implementation

The NAO high anthropomorphism robot functioned autonomously during the experiment, requiring no input or intervention from the experimenter. The robot's verbal and non-verbal behaviors, audio recording, and graphical display for the question prompts on the monitor were controlled through Python scripts initiated at the beginning of the experiment. While the participant was delivering their answer to the robot, we recorded the audio of the participant's responses. When the participant finished giving their answer and pressed the robot's foot bumper, the audio recording stopped so that it could be passed to Whisper API for transcription and then provided as a prompt to GPT 3.5 API for a summary that the robot could include in its positive or negative feedback to the participant. The specific prompt we provided GPT 3.5 was: "I asked someone the question: [QUESTION] they responded: [AUDIO RESPONSE]. Reword their response in English, but replace all instances of I with you and don't use filler words."

2.2.4. Experiment 1 participants

In Experiment 1, a total of 172 participants were recruited from a behavioral science museum. Experiment 1 has 76 Male-identifying participants, 88 Female-identifying participants, 4 Non-Binary participants, and 4 participants who preferred not to answer. Participants' ages ranged from 18 to 74 with an average of 33.06 (SD = 12.42). We also verified no early significant differences among the four conditions in age (BF_{10} = 0.18) and gender (BF_{10} = 0.07). Participants' detailed age, gender (M for Male, F for Female, Other for Other Gender), and gender pair (MM for Male-Male, FF for Female-Female, Mix for Mixed Gender) distributions in Experiment 1 are recorded in Table 1.

Table 1
The age, gender, and gender pair distribution of participants in Experiment 1 (High Anthropomorphism Robot).

	Equal Treatment	Unequal Treatment
Positive Feedback	Age: M = 33.48, SD = 12.95 Gender: 21 M, 23 F, 2 Other Pair: 5 MM, 6 FF, 12 Mix	Age: M = 36.13, SD = 13.32 Gender: 20 M, 20 F, 0 Other Pair: 10 MM, 11 FF, 19 Mix
Negative Feedback	Age: M = 32.30, SD = 10.57 Gender: 17 M, 27 F, 2 Other Pair: 4 MM, 8 FF, 11 Mix	Age: M = 30.40, SD = 12.61 Gender: 18 M, 18 F, 4 Other Pair: 10 MM, 11 FF, 19 Mix

2.2.5. Experiment 1 experimental setting

Experiment 1 was performed in a quiet room that contained a rectangular table and two chairs, shown in Fig. 4. On top of the table, we placed the NAO high anthropomorphism robot in the center and a monitor behind the robot to show the question prompts. The two chairs were placed in front of the table, and each chair was also positioned at an angle of 45° towards the robot and the other participant. We also placed two cameras (a webcam on top of the monitor recording the interaction from the front angle and a camera on a tripod recording from the back).

2.2.6. Experiment 1 measures

To assess how the robot's feedback valence and treatment of the participants affected each pair's interpersonal closeness in Experiment 1, we collected participants' responses to a post-experiment questionnaire, observed each pair's decision-making outcomes, and examined their responses to a post-experiment interview.

Interactive IOS Scale for Multiparty Interactions (IIMI) (Zhang, Lin, et al., 2023). To measure participants' perceived interpersonal closeness among themselves, the other participant, and the robot, we used the Interactive IOS Scale for Multiparty Interactions (IIMI) in the post-experiment questionnaire. The IIMI was first introduced in Zhang et al. (Zhang, Lin, et al., 2023) and was designed based on Inclusion of the Other in the Self Scale from Aron et al. (Aron et al., 1992). As shown in Fig. 5, the IIMI asks participants to drag and drop three circles representing themselves, the other participant, and the robot, to illustrate the closeness the participant perceives between them. From measuring the pixel distances between the center of each circle, we could assess the participant's perceived interpersonal closeness between themselves and the other participant (Self-Other Distance), themselves and the robot (Self-Robot Distance), and the other participant and the robot (Other-Robot Distance).

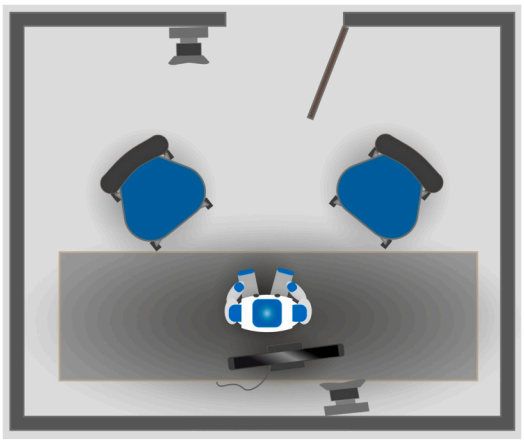

Fig. 4. An illustration of the Experiment 1 setup including two chairs, a table that has a robot and a monitor on top, and two cameras used for video recording.

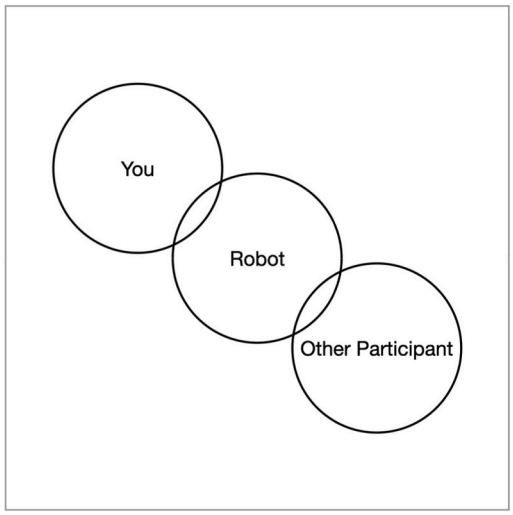

Fig. 5. The Interactive IOS Scale for Multiparty Interactions (IIMI) (Zhang, Lin, et al., 2023): participants dragged three circles into a formation that best represents their interpersonal closeness with the other participant and the robot.

Human-Human Liking Scale (Maxwell et al., 1985). Since closeness and mutual liking are closely related (Aron et al., 1997), we also used the Human-Human Liking Scale (Maxwell et al., 1985) in the post-experiment questionnaire to measure participants' interpersonal closeness. The Human-Human Liking Scale evaluates how much each participant likes the other participant with 14 questions. Each question asks participants to rate a statement (e.g., "I very much enjoyed talking to the other participant.") on a 7-point Likert scale (1 indicates *Strongly Disagree* and 7 indicates *Strongly Agree*).

Robotic Social Attributes Scale (RoSAS) (Carpinella et al., 2017). In the post-experiment questionnaire, we also administered the Robotic Social Attributes Scale (RoSAS) to evaluate participants' perceptions of the robot to capture social dynamics in a human-robot group. The RoSAS contains three subscales: "warmth" (6 questions), "competence" (6 questions), and "discomfort" (6 questions). Each question asks participants to rate a word associated with the perception of the robot such as "Social", "Responsive", and "Awkward" on a 9-point Likert scale (1 indicates *Definitely Not Associated* and 9 indicates *Definitely Associated*).

Video Annotations of Participants' Decision-Making Outcomes. We analyzed how each participant's initial viewpoint on a question prompt was incorporated into the group's final decision. We also computed a measure of how balanced the group's final decisions were between the two participants. Since each participant was placed in one experimental condition dictated by our 2x2 experimental design, this resulted in three possible groupings of participants: (a) Positive-Positive: both participants were in the *positive feedback and equal treatment condition* where the robot responded mostly positively to both participants, (b) Negative-Negative: both participants were in the *negative feedback and equal treatment condition* where the robot responded mostly negatively to both participants, (c) Positive-Negative: one participant was in the *positive feedback and unequal treatment* condition and the other participant was in the *negative feedback and unequal treatment* condition where the robot responded mostly positively to one participant and mostly negatively to the other participant.

To calculate how balanced each group's decisions were between the two participants for each of the eight rounds, we categorized each joint decision as either a_{Shared} (i.e., an answer that is a combination of both participant's initial viewpoint in each round), a_{New} (i.e., an answer that is not related to any participant's initial viewpoint in each round), a_{P1} (i.e., an answer incorporating the initial viewpoint only from the first participant), or a_{P2} (i.e., an answer incorporating the initial viewpoint only from the second participant). We then calculated a balance ratio for each pair of participants using the following formula:

$$Balance\ Ratio = 1 - \left(\frac{abs((a_{P1} + 0.5{*}a_{Shared}) - (a_{P2} + 0.5{*}a_{Shared}))}{a_{P1} + a_{P2} + a_{Shared}}\right)$$

The balance ratio ranges from 0 to 1 with a higher value indicating more balance in incorporating each participant's initial viewpoint to the final decision. For example, if two final decisions originated from the first participant, two final decisions came from the second participant, and four final decisions were shared between both participants' initial answers, then the balance ratio would be 1 showing the maximized balance. For another example, if five final decisions originated from the first participant and the remaining three final decisions were new, then the balance ratio would be calculated to 0, indicating no balance.

Semi-Structured Interview. We conducted a post-experiment semi-structured interview to allow participants to freely express their views while remaining consistent with an interview framework (Galletta, 2013). The semi-structured interviews were conducted in a separate room, where neither the other participant nor the robot were present. The interview included the following questions concerning the overall experience, the other participant, the robot, and the robot's effect on the interaction.

1. Describe your time with the robot and the other participant.
2. Tell me what your impression is of the other participant.
3. What is your impression of the robot?
4. Did the robot influence your interaction with the other participant?

2.2.7. Experiment 1 procedure

After the researcher obtained consent from both participants for joining in the study and video-recording, the participants were brought into the study room. The researcher introduced the robot and told the participants that the robot would ask them to make joint decisions for eight questions and their individual answers need to be different from each other. Then, the researcher pressed the robot's foot bumper to start a tutorial session that familiarized participants with the procedure to interact with the robot such as how and when to press the robot's foot bumper. In the tutorial session, the robot asked the question, "Your friend's in-laws are coming to visit, which meal would you recommend they cook for them?" After each participant gave their individual answer, the robot provided a verbal response that acknowledged and summarized what the participant said without displaying either positive or negative feedback. During the tutorial session, the researcher was present in the room with the participants and would answer any questions from the participants. At the end of the tutorial session, the researcher reminded the participants about the procedure of the study, pressed the robot's foot bumper to start the main part of study, and left the study room.

The robot then mediated the task through the decision-making process for each of the eight progressively deeper and more opinionated discussion questions following the steps in Section 2.1. As soon as the actual study was concluded, the researcher re-entered the room, brought the participants into two separate rooms to fill out the post-experiment questionnaire, and conducted individual semi-structured interviews with each participant about their experience. The total duration of the study was approximately 30 min, and each participant was compensated with 600 points (equivalent to $6.00 USD) for use on prizes at Mindworks.

2.2.8. Experiment 1 data analysis

To analyze the effect of the robot's feedback valence (positive or negative) and treatment (equal or unequal) in Experiment 1, we conducted a quantitative analysis to examine participants' post-experiment questionnaire responses through two-way ANOVA tests with Type III sum of squares using sum contrasts. Before we performed the two-way ANOVA tests, we averaged questions and calculated Cronbach alphas

from the Human-Human Liking Scale ($\alpha = 0.95$), the RoSAS "warmth" subscale ($\alpha = 0.90$), the RoSAS "competence" subscale ($\alpha = 0.91$), and the RoSAS "discomfort" subscale ($\alpha = 0.82$). For the IIMI measure specifically, we excluded participants' data if they did not comply with the prompt to adjust the circles and also if participants experienced difficulty separating circles that became overlapped. For all questionnaire measures, we also excluded outliers in the data (i.e., 3 standard deviations from the mean). In each two-way ANOVA test, we included age and gender as covariates and used an *alpha* level at 0.05 to determine significance. We employed Estimated Marginal Means for post hoc pairwise comparisons. As part of our quantitative analysis for Experiment 1, we also included calculating a balance ratio for how equally likely each participant's initial viewpoint was incorporated into the pair's decision-making outcomes through one-way ANOVA tests. For these one-way ANOVA tests, we employed Estimated Marginal Means for post hoc pairwise comparisons. We verified the inter-rater reliability via Krippendorff's Alpha ($\alpha = 0.88$) by asking the coders to evaluate an overlapping set of 15 videos (17.44 % of all videos).

For the qualitative analysis for Experiment 1, we also conducted a thematic analysis of the semi-structured interviews, which involved the following stages (Boyatzis, 1998; Gibbs, 2008): (1) two coders independently identified initial themes from interview transcriptions and discussed them with a third researcher to form a list of themes, (2) two coders analyzed an overlapping set of 30 participants' transcriptions (17.44 % of all participants' transcriptions) and calculated the overall Cohen's Kappa ($\kappa = 0.86$), and (3) the two coders performed the thematic analysis on the remaining transcriptions. Through this thematic analysis, we identified four themes: overall experience, perception of the other participant, perception of the robot, and perception of the robot's influence on the participant's human-human interaction. We further classified each theme into a set of mutually exclusive labels: *positive*, *negative*, *neutral*, and *not applicable* (cases when participants did not mention anything about the theme).

2.3. Experiment 2 - low anthropomorphism robot

The goal of Experiment 2 was to explore the same set of research questions as Experiment 1 about the influence of a robot's agreement or disagreement on people's interpersonal closeness in a decision-making task, however, with a different robot that exhibited a different level of anthropomorphism. Therefore, the experimental design of Experiment 2 was similar to that of Experiment 1, except that Experiment 2 utilized a lamp-like **low anthropomorphism robot** (Hoffman, Zuckerman, et al., 2015), as shown in Fig. 1B. This robot differed notably in its physical design, which was mechanical and simple, and communication modality that did not involve speech and was based on minimal non-verbal gestures. We intentionally avoided using any audio input (e.g., auditory question prompts) so that participants wouldn't mistake it for the low-anthropomorphism robot's verbal communication. Experiment 2 was conducted on the campus of Reichman University near the city of Tel Aviv, Israel.

2.3.1. Experiment 2 gesture design

We designed the low anthropomorphism robot's gestures via several iterations with an animator and a member of the research team who is a Human-Robot Interaction expert. The design process resulted in the following gestures.

- *Greeting Gesture*: The robot performed a "greeting" gesture by moving up and down (vertically) in a 5 cm amplitude for 12 s.
- *Listening Gesture - Lean and Gaze*: The robot performed a "listening" gesture where it moved towards the participant (either towards its left or right) and up to a position where its top part matched the height of the participant's head.
- *Agreement Gesture - Nodding*: The robot performed movements simulating "agreeing" by nodding its top part towards the participant in a repetitive manner. This gesture had two levels of intensity: (1) Low - 3 nods across 22 s, and (2) High - 8 nods across 22 s. The meaning of the gesture and the intensity were validated in a pilot study.
- *Disagreement Gesture - Short Left Right Shaking*: The robot performed movements simulating "disagreeing" by repeatedly shaking its top part from left to right and back. This gesture had two levels of intensity: (1) Low - 4 shakes across 22 s, and (2) High - 7 shakes across 22 s. The meaning of the gesture and the intensity were validated in a pilot study.
- *Discussion Gesture*: The robot turned to the center and performed slow up and down (vertical) movements in a 3 cm amplitude until the participants reached a joint decision.

These gestures were then sequenced into a fluent robotic behavior that differed for each pair of participants depending on the experimental condition. The understanding of these gestures was validated in a pilot study with 5 pairs (N = 10), in which participants were asked to describe each gesture, its meaning, and their general experience. All participants easily understood the intended meaning of the gestures and reported a clear understanding of the robot's general intention.

2.3.2. Experiment 2 conversation mediation

We designed the robot to mediate the participants' responses through the decision-making process outlined in Section 2.1. After the question prompt was presented on the tablet's screen, the robot would perform the *Listening Gesture* by turning towards one of the participants, indicating that it was their turn to present their initial response. While the participant gave their initial response, the robot performed *Agreement* or *Disagreement* gestures. After 40 s, the robot turned towards the other participant, indicating that it was their turn, and performed gestures consistent with the experimental condition while the second participant provided their initial response. After another 40 s, the robot turned towards the center and performed the *Discussion Gesture*, showing that participants should now perform a discussion and come up with a mutual decision. The order of who spoke first was alternated for every question.

2.3.3. Experiment 2 feedback delivery

We designed the low anthropomorphism robot to perform the relevant non-verbal gestures to express its positive or negative feedback to each participant's individual answer. When responding positively to the participant, the robot performed the *Agreement* gestures while the participant was presenting their ideas. When responding negatively, the robot performed the *Disagreement* gestures. The robot's feedback gestures were presented with increasing intensity, switching from low to high intensity after the fourth question (see Section 2.3.1 for the details of feedback gestures with different intensities).

2.3.4. Experiment 2 technical implementation

During the experiment, the gestures from the low anthropomorphism robot were controlled by a Wizard-of-Oz operator through the Butter Robotics platform (Megidish, 2017). This platform allows researchers to wirelessly execute robotic commands that can navigate the two motors on the robot: a base motor that controls horizontal rotation motion and a body motor that controls vertical motion.

2.3.5. Experiment 2 participants

For Experiment 2, a total of 150 participants were recruited from a university community. Experiment 2 had 44 Male-identifying participants, 106 Female-identifying participants, 0 Non-Binary participants, and 0 participants who preferred not to answer. The age of the participants varied from 18 to 59 with an average of 24.23 ($SD = 7.04$). We verified no significant differences between the four conditions in participants' age ($BF_{10} = 0.12$) and gender ($BF_{10} = 0.06$). Participants' detailed age, gender (M for Male, F for Female, Other for Other Gender),

and gender pair (MM for Male-Male, FF for Female-Female, Mix for Mixed Gender) distributions in Experiment 2 are recorded in Table 2.

2.3.6. Experiment 2 experimental setting

Similar to Experiment 1, Experiment 2 was also conducted in a quiet room at a research lab. The setup included a table (70 cm in height) with a tablet to show the question prompts and the robot placed in the center, and its height reached that of a human's shoulder when seated. The two chairs were placed 76-cm apart since it is considered to be a comfortable conversation distance (Burgoon and Hale, 1984). A small camera was used for recording via the Zoom platform on the wall behind the robot (see Fig. 6).

2.3.7. Experiment 2 measures

The measures for quantitative analysis in Experiment 2 were the same as those in Experiment 1 (see Section 2.2.6). The semi-structured interview included the following questions.

1. Describe the experience.
2. How did you feel during the experience?
3. Describe the other participant.
4. Describe the robot.
5. What did you think about your performance?
6. Which factors influenced your performance?

2.3.8. Experiment 2 procedure

A few days before the experiment, participants received two questionnaires by email: an Attitude Towards Robots questionnaire (Nomura et al., 2006) and a demographic questionnaire to verify that the groups were balanced by gender pair. When participants arrived at the lab, the researcher assessed their level of previous acquaintance and told them they would be recorded. All participants signed a consent form and were informed that the recorded videos would be erased after data analysis. The researcher explained that the participants would be asked to make joint decisions regarding eight question prompts presented by a tablet screen with the presence of a robotic object. The researcher further explained that each of the participants would get 40 s to state their answers before having a discussion to reach a joint decision. Participants were also informed that a robotic object would be present in the room and would mediate the timing and turn-taking for participants to state their answers and reach a joint decision. They were also told that each of them could talk when the robot turned in their direction and that they should discuss their answers when the robot turned toward the center. Lastly, participants were told that one of them had to clearly state the joint decision once they finished their discussion.

The researcher then escorted the participants into the experiment room and directed each participant toward one of the two seats. The robot performed the *Greeting Gesture* as they took their seats. The experiment began with a tutorial session of one question ("Your friend's in-laws are coming to visit, which meal would you recommend they cook for them?") and the robot's *Listening Gesture* and *Discussion Gesture*. After verifying that the instructions were clear, the researcher left the room and remotely activated the presentation of the question prompts on the tablet and the robotic behavior relevant to the experimental condition. For each question, the robot performed *Listening Gesture* showing which participant should speak, *Agreement Gesture* and *Disagreement Gesture* showing the feedback valence, and *Discussion Gesture* indicating that participants should reach a joint decision. After discussing the eight questions, each participant was escorted by the researcher to a separate room where they filled out questionnaires and participated in a semi-structured interview. At the final stage of the experiment, participants were asked to describe a recent positive experience (to mitigate any negative effects). The researcher debriefed the participants and verified that they left with an overall positive experience. Each participant was compensated with extra course credits or a gift card (equivalent to $15.00 USD) for local shops.

Table 2
The age, gender, and gender pair distribution of participants in Experiment 2 (Low Anthropomorphism Robot).

	Equal Treatment	Unequal Treatment
Positive Feedback	Age: $M = 25.05$, $SD = 9.48$ Gender: 12 M, 28 F, 0 Other Pair: 2 MM, 10 FF, 8 Mix	Age: $M = 25.28$, $SD = 8.50$ Gender: 12 M, 23 F, 0 Other Pair: 4 MM, 19 FF, 12 Mix
Negative Feedback	Age: $M = 22.75$, $SD = 2.57$ Gender: 12 M, 28 F, 0 Other Pair: 2 MM, 10 FF, 8 Mix	Age: $M = 23.94$, $SD = 5.42$ Gender: 8 M, 27 F, 0 Other Pair: 4 MM, 19 FF, 12 Mix

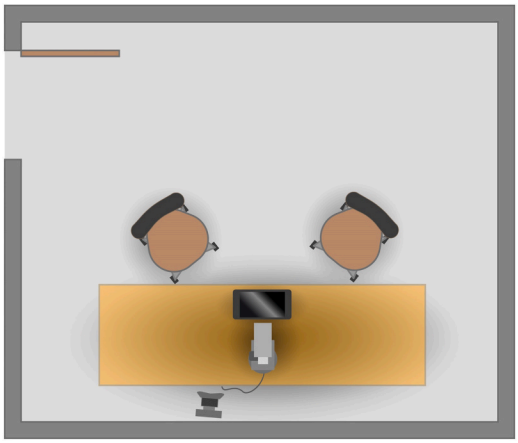

Fig. 6. An illustration of the Experiment 2 setup showing two chairs, a table with the robotic object and a tablet on top, and one camera placed on the wall for recording.

2.3.9. Experiment 2 data analysis

For the quantitative analysis, Experiment 2 employed the same two-way ANOVA tests detailed in Experiment 1 to examine the influence of the valence of the robot's feedback (*positive* or *negative*) and the robot's treatment (*equal* or *unequal*) on each participant's questionnaire responses. In Experiment 2, we also averaged questions and calculated Cronbach alphas from the Human-Human Liking Scale ($\alpha = 0.88$), the RoSAS "warmth" subscale ($\alpha = 0.89$), the RoSAS "competence" subscale ($\alpha = 0.88$), and the RoSAS "discomfort" subscale ($\alpha = 0.76$). The quantitative analysis for Experiment 2 also used the same one-way ANOVA tests to examine the balance ratio of the pair's decision-making outcomes.

For the qualitative analysis, Experiment 2 also included a thematic analysis of the interviews with the same stages described in Experiment 1 (see Section 2.2.8). The coders reviewed an overlapping set of interview transcriptions (20 % of all participants' transcriptions; $\kappa = 0.87$). The interview transcriptions were categorized into four themes: the robot's agreement or disagreement with the participant, perception of the other participant, perception of the robot, and perception of the robot's influence on the participant's human-human interaction. For the theme of the robot's agreement or disagreement with the participant, we further classified it into a set of mutually exclusive labels: *agreed with me more*, *agreed with the other more*, or *agreed with both equally*. For the remaining themes, we classified them into three mutually exclusive labels: *positive*, *negative*, and *neutral*.

3. Results

For each ANOVA test in quantitative analysis, we report the F-statistic (F) and the effect size as partial eta squared (η_p^2). For the qualitative

analysis, we report the counts and percentages of each theme in the thematic analysis. The anonymized data from Experiment 1 and Experiment 2 can be accessed through this OSF link.[3]

3.1. Experiment 1 (high anthropomorphism robot)

The quantitative data analysis in Experiment 1, including means, standard deviations, F statistics, effect sizes, and p values of each measure, is summarized in Tables 3 and 4.

3.1.1. Interpersonal distance perceived between interactants

To assess the robot's impact on the interpersonal closeness between the two participants, we examined their responses to the Interactive IOS Scale for Multiparty Interactions (IIMI) (Zhang, Lin, et al., 2023). We display the averaged participant responses to this scale based on the four conditions in Fig. 7.

Interpersonal Distance between the Two Participants (Self-Other Distance): We first examined the interpersonal distance participants reported between themselves and the other participant (Self-Other Distance). We found that the valence of the robot's feedback had a significant influence on the participants' reported Self-Other Distance ($F(1,146) = 4.62$, $\eta_p^2 = 0.03$, $p = 0.033$). Participants who experienced *positive feedback* from the robot reported a closer Self-Other Distance ($M = 89.08$ pixels, $SD = 69.77$) compared to those who experienced *negative feedback* from the robot ($M = 114.14$ pixels, $SD = 64.14$) shown in Fig. 8A. This finding supports H_{2a} (Influence of Negative Affect), which states that participants perceive greater interpersonal closeness with the other participant when the robot gives them positive feedback as opposed to negative feedback. In contrast, this finding does not support H_{1a} (Heider's Balance Theory), which states that the robot's equal treatment, as opposed to the robot's unequal treatment, towards two participants will make them perceive greater interpersonal closeness with each other. We did not observe the robot's treatment or the interaction between the robot's feedback valence and its treatment significantly impacted participants' reported Self-Other Distance.

Interpersonal Distance between the Participant and the Robot (Self-Robot Distance): We next analyzed the interpersonal distance participants reported between themselves and the robot (Self-Robot Distance). We observed that the robot's feedback valence significantly affected the participants' reported Self-Robot Distance ($F(1,155) = 21.01$, $\eta_p^2 = 0.12$, $p < 0.001$). Participants who received *positive feedback* from the robot rated the Self-Robot Distance significantly closer ($M = 99.87$ pixels, $SD = 58.33$) than those who received *negative feedback* ($M = 139.20$ pixels, $SD = 48.98$; see Fig. 8B). We saw no effects of the robot's treatment or the interaction between the robot's valence of feedback and its treatment on participants' reported Self-Robot Distance.

Interpersonal Distance between the Other Participant and the Robot (Other-Robot Distance): For the distance between the other participant and the robot (Other-Robot Distance), we did not find a significant impact of either the robot's feedback valence or the robot's treatment alone. However, we did find a statistically significant interaction between the valence of the robot's feedback and the robot's treatment ($F(1, 153) = 28.46$, $\eta_p^2 = 0.16$, $p < 0.001$; see Fig. 8C). In *positive feedback* conditions, participants' estimation of Other-Robot Distance was dependent on the robot's treatment. Specifically, participants in the *positive feedback and equal treatment* condition ($M = 96.89$ pixels, $SD = 47.62$) reported a significantly closer Other-Robot Distance compared to those in *positive feedback and unequal treatment* condition ($M = 128.52$ pixels, $SD = 48.97$, $p = 0.026$) where the other participant was treated negatively by the robot. In *negative feedback* conditions, participants' ratings of Other-Robot Distance depended on the robot's treatment in an opposite pattern. Participants in *negative feedback and equal treatment* condition ($M = 143.15$ pixels, $SD = 40.11$) reported a significantly greater Other-Robot Distance compared to participants in the *negative feedback and unequal treatment* condition ($M = 91.06$ pixels, $SD = 54.48$, $p < 0.001$) where the other participant was treated positively by the robot. At the same time, in *equal treatment* conditions, participants' ratings of Other-Robot Distance were influenced by the robot's feedback valence. In particular, participants in the *positive feedback and equal treatment* condition ($M = 96.89$ pixels, $SD = 47.62$) rated a significantly closer Other-Robot Distance than those in the *negative feedback and equal treatment* condition ($M = 143.15$ pixels, $SD = 40.11$, $p < 0.001$). In *unequal treatment* conditions, participants' assessments of Other-Robot Distance from participants were also affected by the robot's feedback valence. Participants in the *negative feedback and unequal treatment* condition ($M = 91.06$ pixels, $SD = 54.48$) rated a significantly closer Other-Robot Distance than those in the *positive feedback and unequal treatment* condition ($M = 128.52$ pixels, $SD = 48.97$, $p = 0.009$).

3.1.2. Human-human liking between the two participants

To further assess participants' interpersonal closeness, we examined participants' responses to the Human-Human Liking Scale (Maxwell et al., 1985), which indicates how much each participant liked the participant they were paired with. We did not find a significant influence of either the robot's feedback valence or the robot's treatment alone. Thus, in this measure, we did not find any evidence supporting H_{1a} (Heider's Balance Theory), where the robot's equal treatment as opposed to unequal treatment of the participants would lead participants to perceive greater interpersonal closeness between them. We also did not find support to H_{2a} (Influence of Negative effect), where participants would perceive greater interpersonal closeness with each other when the robot shows positive feedback rather than negative feedback. Nevertheless, we did find a statistically significant interaction between the robot's feedback valence and the robot's treatment on how much participants reported liking the other human participant ($F(1,163) = 3.94$, $\eta_p^2 = 0.02$, $p = 0.049$). Post hoc pairwise comparisons indicated that the impact of the feedback's valence depended on the treatment. In the *equal treatment* conditions, participants who experienced positive feedback from the robot expressed a significantly greater liking for the other participant ($M = 5.73$, $SD = 0.85$) than those who experienced *negative feedback* ($M = 5.25$, $SD = 1.05$, $p = 0.035$). In the *unequal treatment* conditions, the robot's feedback valence had no impact on the participants' liking ratings (see Fig. 9). The other pairwise comparisons were not statistically significant.

3.1.3. Participants' perceptions of the robot's social attributes

We examined how the valence of the robot's feedback and the robot's treatment affected the participant's perception of the robot's social attributes (warmth, competence, discomfort), using the participant's ratings on the Robotic Social Attributes Scale (RoSAS) (Carpinella et al., 2017).

We found the robot's feedback valence to have a significant influence on how warm participants perceived the robot to be ($F(1,165) = 13.22$, $\eta_p^2 = 0.07$, $p < 0.001$; see Fig. 10A). Participants who experienced *positive feedback* from the robot felt the robot was significantly warmer ($M = 4.80$, $SD = 1.74$) compared to those who received *negative feedback* from the robot ($M = 3.85$, $SD = 1.91$). We found that neither the robot's treatment nor the interaction between the robot's feedback valence and treatment had an important role in how participants rated the robot's warmth.

We also saw that the valence of feedback from the robot has a significant effect on how competent participants perceived the robot to be ($F(1,165) = 8.61$, $\eta_p^2 = 0.05$, $p = 0.004$; see Fig. 10B). Those who received *positive feedback* considered the robot significantly more competent ($M = 6.05$, $SD = 1.82$) compared to those who received *negative feedback* ($M = 5.33$, $SD = 1.83$). The robot's treatment or the interaction between the robot's feedback valence and treatment did not

[3] https://osf.io/wc9nj/?view_only=7425e31cb61746d09478c29024a6baf5.

Table 3
Means, standard deviations, F statistics, effect sizes of the measures, and P values across the two main variables in experiment 1.

	Positive Feedback		Negative Feedback		F	η_p^2	p
	M	SD	M	SD			
Self-Other Distance	89.08	69.77	114.14	64.14	$F(1,146) = 4.62$	0.03	0.033(*)
Self-Robot Distance	99.87	58.33	139.20	48.98	$F(1,155) = 21.01$	0.12	<0.001(***)
Other-Robot Distance	111.89	50.51	119.01	53.79	$F(1,153) = 0.35$	0.00	0.554
Human-Human Liking	5.61	0.86	5.37	0.98	$F(1,163) = 2.81$	0.02	0.095
RoSAS Warmth	4.80	1.74	3.85	1.91	$F(1,165) = 13.22$	0.07	<0.001(***)
RoSAS Competence	6.05	1.82	5.33	1.83	$F(1,165) = 8.61$	0.05	0.004(**)
RoSAS Discomfort	2.51	1.21	2.76	1.28	$F(1,161) = 1.83$	0.01	0.177
	Equal Treatment		Unequal Treatment		F	η_p^2	p
	M	SD	M	SD			
Self-Other Distance	104.57	70.16	98.36	65.96	$F(1,146) = 0.27$	0.00	0.603
Self-Robot Distance	117.50	51.83	122.35	62.87	$F(1,155) = 0.51$	0.00	0.474
Other-Robot Distance	120.84	49.44	109.54	54.83	$F(1,153) = 1.70$	0.01	0.195
Human-Human Liking	5.49	0.98	5.49	0.87	$F(1,163) = 0.03$	0.00	0.853
RoSAS Warmth	4.47	2.03	4.15	1.70	$F(1,165) = 1.20$	0.01	0.275
RoSAS Competence	5.89	1.96	5.45	1.71	$F(1,165) = 2.26$	0.01	0.134
RoSAS Discomfort	2.39	1.18	2.92	1.27	$F(1,161) = 7.14$	0.04	0.008(**)

(*), (**), (***) denote $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively.

Table 4
Means, standard deviations, F statistics, effect sizes of the measures, and P values across the four conditions in experiment 1.

	Positive Equal		Positive Unequal		Negative Equal		Negative Unequal		F	η_p^2	p
	M	SD	M	SD	M	SD	M	SD			
Self-Other Distance	88.68	68.62	89.51	71.93	121.30	68.71	106.98	59.27	$F(1,146) = 0.43$	0.00	0.513
Self-Robot Distance	93.88	46.33	106.49	69.28	140.05	46.89	138.21	51.92	$F(1,155) = 0.45$	0.00	0.503
Other-Robot Distance	96.89	47.62	128.52	48.97	143.15	40.11	91.06	54.48	$F(1,153) = 28.46$	0.16	< 0.001(***)
Human-Human Liking	5.73	0.85	5.46	0.86	5.25	1.05	5.51	0.88	$F(1,163) = 3.94$	0.02	0.049
RoSAS Warmth	5.10	1.89	4.45	1.51	3.84	1.99	3.85	1.84	$F(1,165) = 0.98$	0.01	0.325
RoSAS Competence	6.29	1.87	5.77	1.75	5.49	1.99	5.14	1.63	$F(1,165) = 0.00$	0.00	0.945
RoSAS Discomfort	2.14	1.15	2.94	1.13	2.65	1.16	2.89	1.41	$F(1,161) = 2.05$	0.01	0.154

(***) denotes $p < 0.001$.

have a significant influence on participants' ratings of the robot's competence.

We found the robot's treatment of the two participants played a significant role in how much discomfort participants reported feeling from the robot ($F(1,161) = 7.14$, $\eta_p^2 = 0.04$, $p = 0.008$; see Fig. 10C). Participants who experienced *unequal treatment* from the robot rated the robot as causing more discomfort ($M = 2.92$, $SD = 1.27$), compared to those who experienced *equal treatment* ($M = 2.39$, $SD = 1.18$). However, we did not see the robot's feedback valence or the interaction between the robot's feedback valence and treatment significantly affecting participants' thoughts of the robot's discomfort level.

From the RoSAS "warmth" and "competence" subscales, we found evidence to support H_{2c} (Influence of Negative Affect), where participants perceive the robot as warmer, more competent, and less discomforting when the robot gives positive feedback rather than negative feedback. However, from the RoSAS "discomfort" subscale, we did not find support for H_{2c} (Influence of Negative Affect).

3.1.4. Decision-making outcomes

We did not find a significant main effect from the three types of pairs (i.e., Positive-Positive, Negative-Negative, Positive-Negative) on the balance ratio, which calculates how balanced each participant's initial viewpoints are incorporated into the group's joint decision. Thus, we found no support for H_{1b} (Heider's Balance Theory), where each participant's viewpoints will be equally likely to be incorporated into the group's final decision when the robot gives equal treatment as opposed to unequal treatment. We also found no evidence for H_{2b} (Influence of Negative Affect), where the robot's positive feedback as opposed to negative feedback will make each participant's viewpoints equally likely to be integrated into the final decision of the group.

3.1.5. Thematic analysis

In Experiment 1, the thematic analysis resulted in four themes: overall experience, perception of the other participant, perception of the robot, and perception of the robot's influence on the participant's human-human interaction.

Theme 1: Overall Experience. 144 out of 172 participants (84 %) discussed their overall experience during the interview. Participants had overwhelmingly positive experiences across all conditions. However, slightly more participants spoke positively about their overall experience when the robot gave positive feedback (*equal treatment* (38/42, 90 %) and *unequal treatment* (29/33, 88 %)) in comparison to negative feedback (*equal treatment* (29/39, 74 %) and *unequal treatment* (25/30, 83 %)). Participants discussed positive aspects of their interlocutor: *"It was fun chatting with her"* (p.93, *positive feedback, unequal treatment*). They also made positive remarks about the experience in general: *"It was very cool"* (p.29, *negative feedback, unequal treatment*).

Theme 2: Perception of the Other Participant. 170 out of 172 participants (99 %) mentioned their thoughts on the other participant. Of those 170 participants, 149 (88 %) stated having an overall positive perception of the other participant. The distribution of the positive perception of the other participant was similar across the conditions. Participants discussed positive aspects of the other participant: "*[He was] a very good collaborator and a good listener*" (p.22, *positive feedback, equal treatment*); "*He seemed really sociable and very committed*" (p.55, *negative feedback, unequal treatment*).

Theme 3: Perception of the Robot. From the 155 out of 172 participants (90 %) who commented on their perceptions of the robot, their

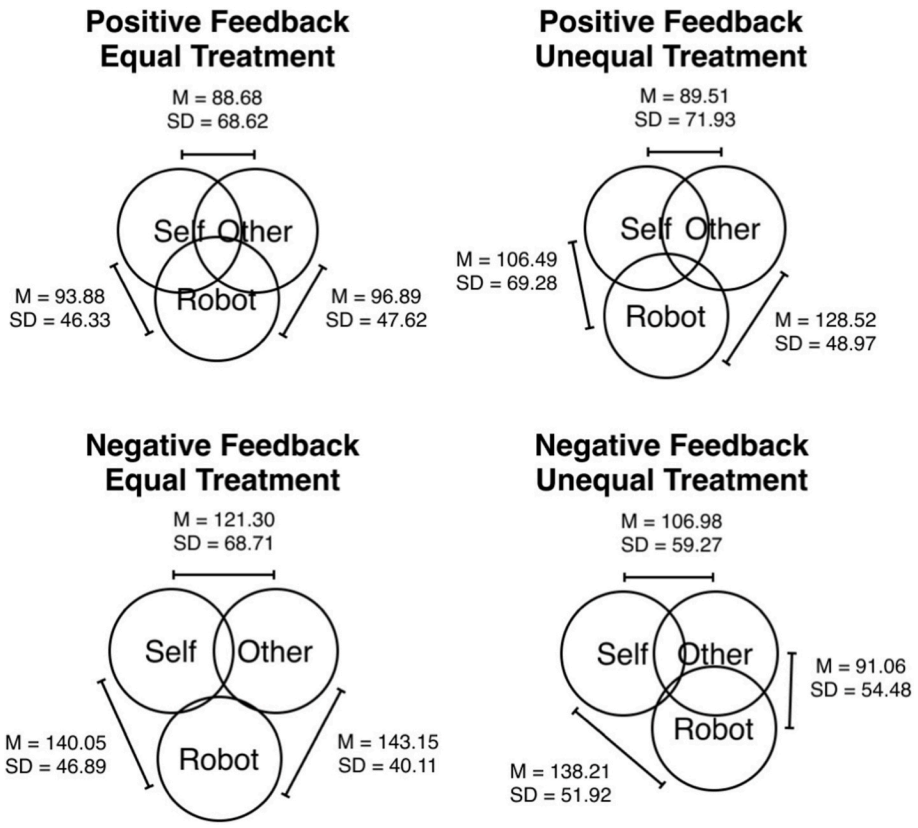

Fig. 7. We display Experiment 1 participants' averaged response to the IIMI scale in each combination of the robot's feedback and treatment type. Participants reported a significantly closer interpersonal distance between them ("Self") and the other human participant ("Other") when the high anthropomorphism robot expressed *positive feedback* (top row) as opposed to *negative feedback* (bottom row).

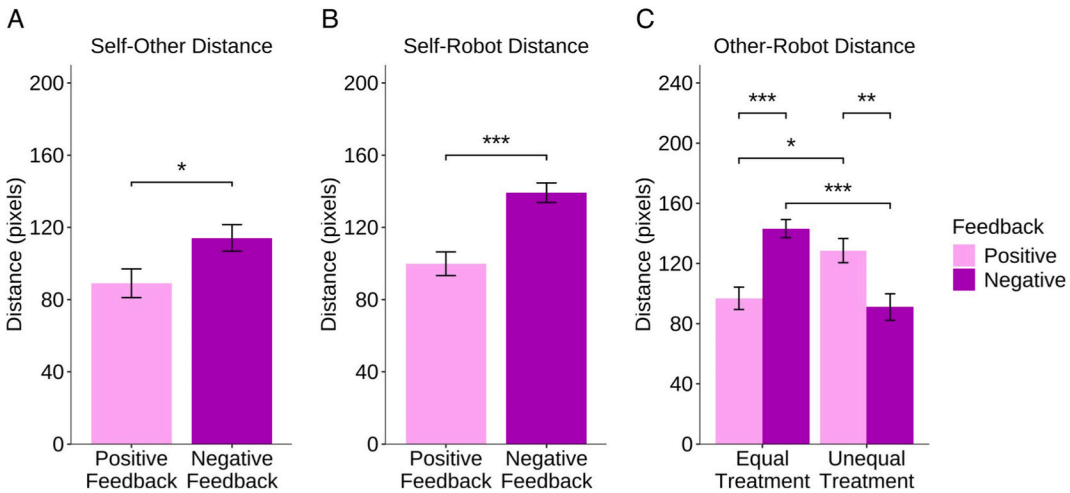

Fig. 8. In Experiment 1, when the high anthropomorphism robot expressed *positive feedback* as opposed to *negative feedback*, participants reported a significantly closer (a) Self-Other Distance and (b) Self-Robot Distance. (c) The interaction between the robot's feedback valence and the robot's treatment has a significant influence on participants' responses to Other-Robot Distance. Error bars show one standard error from the mean. (*), (**), (***) denote $p < 0.05$, $p < 0.01$, and $p < 0.001$ respectively.

responses varied across the four conditions. The *positive feedback and equal treatment* condition (26/43, 60 %) was the only condition where most participants mentioned positive perceptions of the robot: "*I thought [the robot] tried to empathize with us quite a bit and validate us, which probably led us feeling more comfortable and sharing more*" (p.35, *positive feedback, equal treatment*). In the *negative feedback* conditions, *equal treatment* condition (12/41, 29 %) and *unequal treatment* condition (12/37, 32 %), positive perception of the robot appeared less frequently. Participants described the robot as "*opinionated*" (p.5, *negative feedback, equal treatment*) and "*trying to…get you guys to argue*" (p.38, *negative feedback, equal treatment*). Participants in *positive feedback and unequal treatment* condition (12/34, 35 %) also tended to report positive perceptions of the robot less frequently, as some of them mentioned: "*[The robot] was programmed to make us split*" (p.66, *positive feedback, unequal treatment*).

Theme 4: Perception of the Robot's Influence on the Human-

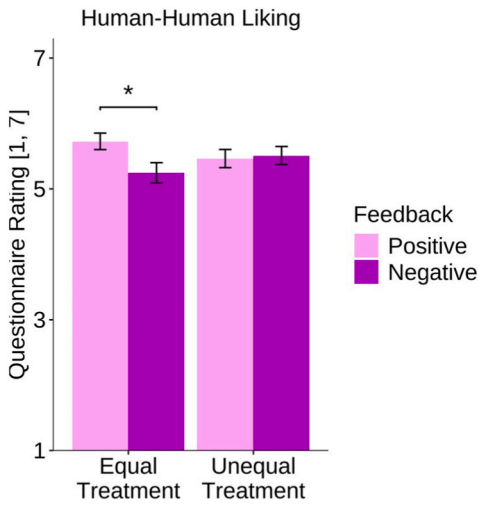

Fig. 9. In Experiment 1, participants in the *positive feedback and equal treatment* condition rated significantly higher liking than those in the *negative feedback and equal treatment* condition. Error bars show one standard error from the mean. (*) denotes $p < 0.05$.

Human Interaction. 107 out of 172 participants (62 %) discussed their opinions on how the robot affected their interaction with the other participant. Interestingly, most participants in all conditions (77/107, 72 %) described the robot as a **neutral** mediator: "*the robot wasn't an active participant in our discussions and was more just there to ask us questions and provide feedback*" (p.27, *negative feedback, equal treatment*). Among those who did not perceive the robot as neutral, most participants described its impact as positive. This effect was slightly more dominant in the *negative feedback and equal treatment* condition (11/30, 37 %) in comparison to the other conditions: *positive feedback and equal treatment* (3/27, 11 %), *positive feedback and unequal treatment* (6/23, 26 %), and *negative feedback and unequal treatment* (6/27, 22 %). Participants in the *negative feedback and equal treatment* condition mentioned bonding over the robot's negativity and responding with humor: "*It actually made us forming an alliance to say this is how we're going to deal with the robot*" (p.39, *negative feedback, equal treatment*); "*When the robot like started disagreeing so quickly, and I think it's kind of like, we kind of became a team a little bit more*" (p.23, *negative feedback, equal treatment*).

3.2. Experiment 2 (low anthropomorphism robot)

The quantitative data analysis in Experiment 2, including means, standard deviations, F statistics, effect sizes, and p values of each measure, is summarized in Tables 5 and 6.

3.2.1. Interpersonal distance perceived between interactants

Similar to Experiment 1, we analyzed participants' responses to the Interactive IOS Scale for Multiparty Interactions (IIMI) (Zhang, Lin, et al., 2023) to assess participants' perceived interpersonal closeness with the other human participant. We display the averaged participant responses to this scale based on the four conditions in Fig. 11.

Interpersonal Distance between the Two Participants (Self-Other Distance): We found the robot's treatment of the two participants had a significant effect on the participants' reported Self-Other Distance ($F(1,141) = 6.13$, $\eta_p^2 = 0.04$, $p = 0.014$). Participants who experienced *equal treatment* from the robot reported a significantly closer Self-Other Distance ($M = 116.02$ pixels, $SD = 62.48$) in comparison to participants who experienced *unequal treatment* from the robot ($M = 140.25$ pixels, $SD = 57.73$; see Fig. 12A). This finding supports H_{1a} (Heider's Balance Theory), which states that participants perceive greater interpersonal closeness when a robot exhibits equal treatment as opposed to unequal treatment. We also saw that the robot's feedback valence had a significant influence on how participants rated Self-Other Distance ($F(1,141) = 4.01$, $\eta_p^2 = 0.03$, $p = 0.047$). Participants who received the robot's *positive feedback* reported a larger Self-Other Distance ($M = 136.44$ pixels, $SD = 71.06$) compared to participants who received *negative feedback* ($M = 117.89$ pixels, $SD = 48.31$; see Fig. 12B). We thus found no evidence to support H_{2a} (Influence of Negative Affect) which states that a robot's positive feedback would make participants perceive greater interpersonal closeness with one another compared to the robot's negative feedback. The interaction between the valence of the robot's feedback and the treatment of the robot did not significantly influence the participants' reported Self-Other Distance.

Interpersonal Distance between the Participant and the Robot (Self-Robot Distance): The robot's feedback valence significantly affected participants' assessment for Self-Robot Distance ($F(1,142) = 10.01$, $\eta_p^2 = 0.07$, $p = 0.002$). Participants who had the robot's *positive feedback* reported a

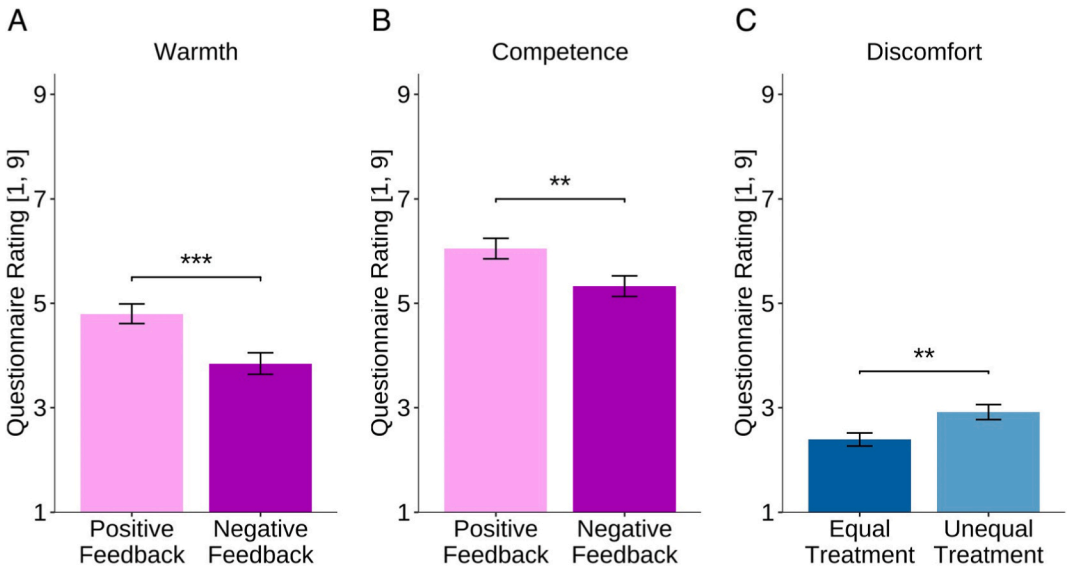

Fig. 10. In Experiment 1, participants rated the high anthropomorphism robot as significantly (a) warmer and (b) more competent when the robot gave them *positive feedback* as opposed to *negative feedback*. (c) Participants also rated the high anthropomorphism robot as significantly more discomforting when it gave *unequal treatment* compared to when it gave *equal treatment*. Error bars show one standard error from the mean. (**) denotes $p < 0.01$ and (***) denotes $p < 0.001$.

Table 5
Means, standard deviations, F statistics, effect sizes of the measures, and P values across the two main variables in experiment 2.

	Positive Feedback		Negative Feedback		F	η_p^2	p
	M	SD	M	SD			
Self-Other Distance	136.44	71.06	117.89	48.31	F(1,141) = 4.01	0.03	0.047(*)
Self-Robot Distance	118.29	45.78	142.44	47.15	F(1,142) = 10.01	0.07	0.002(**)
Other-Robot Distance	128.23	47.35	134.77	53.44	F(1,144) = 0.29	0.00	0.590
Human-Human Liking	5.34	0.77	5.16	0.78	F(1,143) = 2.35	0.02	0.128
RoSAS Warmth	3.76	1.90	3.46	1.90	F(1,144) = 0.93	0.01	0.337
RoSAS Competence	5.30	1.92	5.01	1.95	F(1,144) = 1.04	0.01	0.310
RoSAS Discomfort	3.30	1.37	3.44	1.39	F(1 143) = 0.09	0.00	0.771
	Equal Treatment		Unequal Treatment		F	η_p^2	p
	M	SD	M	SD			
Self-Other Distance	116.02	62.48	140.25	57.73	F(1,141) = 6.13	0.04	0.014(*)
Self-Robot Distance	128.03	46.10	132.97	49.98	F(1,142) = 0.47	0.00	0.493
Other-Robot Distance	131.76	52.84	131.21	47.90	F(1,144) = 0.00	0.00	0.973
Human-Human Liking	5.14	0.83	5.37	0.69	F(1,143) = 3.44	0.02	0.066
RoSAS Warmth	3.43	1.89	3.81	1.91	F(1,144) = 1.45	0.01	0.231
RoSAS Competence	5.23	1.92	5.06	1.96	F(1,144) = 0.27	0.00	0.603
RoSAS Discomfort	3.32	1.43	3.43	1.32	F(1,143) = 0.29	0.00	0.588

(*) and (**) denote $p < 0.05$ and $p < 0.01$, respectively.

Table 6
Means, standard deviations, F statistics, effect sizes of the measures, and P values across the four conditions in experiment 2.

	Positive Equal		Positive Unequal		Negative Equal		Negative Unequal		F	η_p^2	p
	M	SD	M	SD	M	SD	M	SD			
Self-Other Distance	129.13	76.01	145.05	64.81	102.57	41.45	135.46	50.19	F(1,141) = 0.66	0.00	0.417
Self-Robot Distance	122.73	44.32	113.34	47.51	133.32	47.79	152.60	44.94	F(1,142) = 2.65	0.02	0.106
Other-Robot Distance	127.39	48.24	129.20	46.99	136.13	57.35	133.22	49.39	F(1,144) = 0.18	0.00	0.674
Human-Human Liking	5.27	0.76	5.42	0.78	5.00	0.89	5.33	0.60	F(1,143) = 0.22	0.00	0.637
RoSAS Warmth	3.42	2.02	4.14	1.70	3.44	1.77	3.48	2.08	F(1,144) = 1.18	0.01	0.280
RoSAS Competence	5.49	2.14	5.07	1.63	4.97	1.65	5.06	2.27	F(1,144) = 0.41	0.00	0.524
RoSAS Discomfort	3.01	1.34	3.62	1.34	3.62	1.47	3.23	1.29	F(1,143) = 5.02	0.03	0.026(*)

(*) denotes $p < 0.05$.

significantly closer Self-Robot Distance ($M = 118.29$ pixels, $SD = 45.78$) than those who had the robot's *negative feedback* ($M = 142.44$ pixels, $SD = 47.16$; see Fig. 12C). The robot's treatment and the interaction between the robot's treatment and feedback valence had no direct effect on the participants' reported Self-Robot Distance.

Interpersonal Distance Between the Other Participant and the Robot (Other-Robot Distance): We did not find any significant effects of the robot's feedback valence or treatment of the two participants on participants' reported Other-Robot Distance.

3.2.2. Human-human liking between the two participants

When examining participants' responses to the Human-Human Liking Scale (Maxwell et al., 1985), we did not find any significant effects of the robot's feedback valence or treatment of the two participants on how much they reported liking the other participant. We thus did not find support for H_{1a} (Heider's Balance Theory) or H_{2a} (Influence of Negative Affect).

3.2.3. Participants' perceptions of the robot's social attributes

From analyzing participants' responses to the Robot Social Attributes Scale (RoSAS) (Carpinella et al., 2017), we found that the interaction between the robot's feedback valence and treatment had a significant effect on how participants rated their discomfort with the robot ($F(1, 143) = 5.02$, $\eta_p^2 = 0.03$, $p = 0.027$). However, pairwise comparisons did not show any significant differences between individual conditions. Neither the robot's feedback valence nor its treatment showed any significant effect on the "discomfort" subscale. For perceptions of the robot's warmth and competence, we did not find any significant influence from either the robot's feedback valence or the robot's treatment of the two participants. Therefore, we did not find support for H_{2c} (Influence of Negative Affect).

3.2.4. Decision-making outcomes

We excluded one Positive-Positive pair's decision-making outcomes due to a corrupted video file. After excluding this data point, we did not find that the three types of pairs (i.e., Positive-Positive, Negative-Negative, Positive-Negative) had a significant main effect on the balance ratio, which indicates how balanced each participant's initial viewpoints are incorporated into the group's joint decision. Hence, we did not find support for either H_{1b} (Heider's Balance Theory) or H_{2b} (Influence of Negative Affect).

3.2.5. Thematic analysis

In Experiment 2, interview transcriptions were categorized into four themes: the robot's agreement or disagreement with the participant, perception of the other participant, perception of the robot, and perception of the robot's influence on the participant's human-human interaction.

Theme 1: The Robot's Agreement or Disagreement with the Participant. 81 out of 150 participants (54 %) explicitly discussed the robot's agreement or disagreement during their interview responses. In both *equal treatment* conditions, *positive feedback* (14/14, 100 %) and *negative feedback* (16/20, 80 %), almost all participants who reported this theme stated that the robot agreed with both equally: *"We got the same responses from the robot"* (p.25, *positive feedback, equal treatment*); *"[The robot] disagreed with both of us"* (p.89, *negative feedback, equal treatment*). In the *unequal treatment* conditions, participants' responses depended on the feedback valence. From the *negative feedback* condition

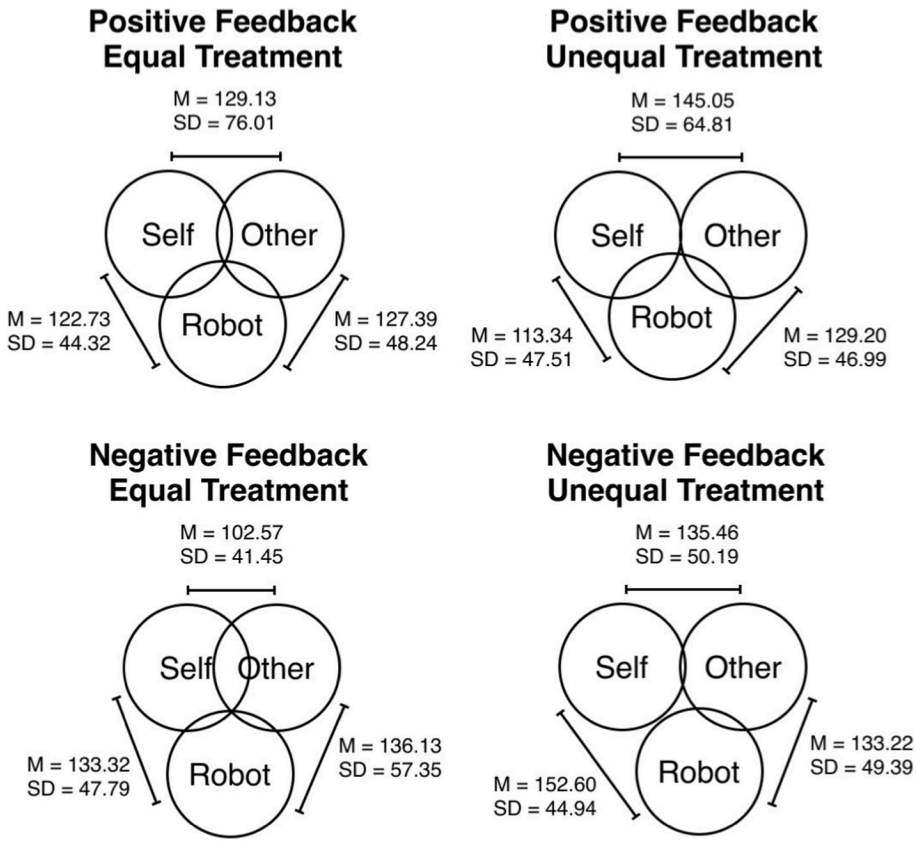

Fig. 11. We show Experiment 2 participants' averaged response to the IIMI scale in each combination of feedback and treatment type. Participants reported a significantly closer interpersonal distance between them ("Self") and the other human participant ("Other") when the low anthropomorphism robot expressed *equal treatment* (left column) as opposed to *unequal treatment* (right column) and when the low anthropomorphism robot expressed *negative feedback* (bottom row) as opposed to *positive feedback* (top row).

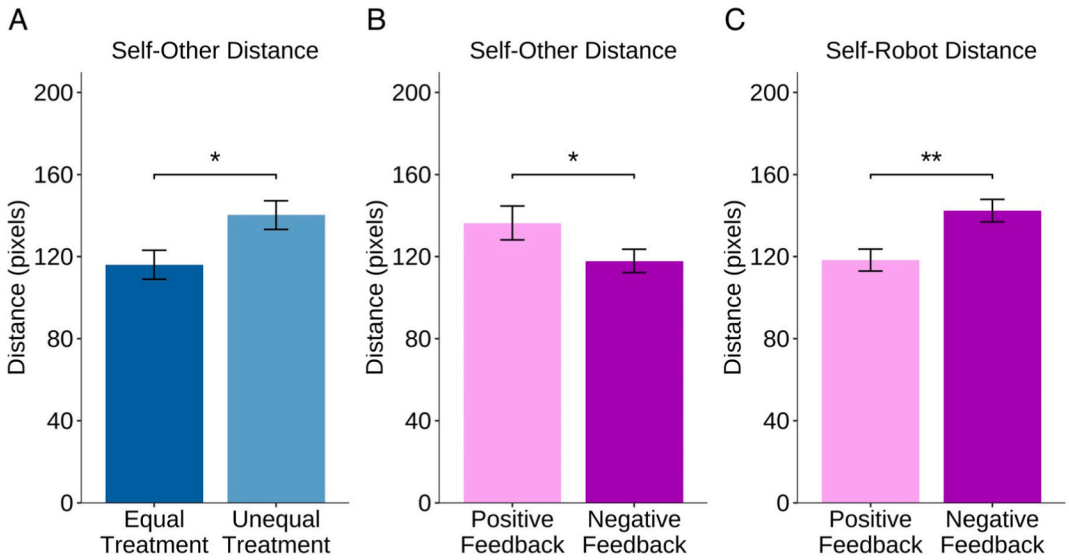

Fig. 12. In Experiment 2, participants rated Self-Other Distance significantly closer when the robot gave (a) *equal treatment* as opposed to *unequal treatment* (b) *negative feedback* as opposed to *positive feedback*. (c) They also rated the Self-Robot Distance significantly closer when the robot gave *positive feedback* instead of *negative feedback*. Error bars show one standard error from the mean. (*) denotes $p < 0.05$ and (**) denotes $p < 0.01$.

(21/23, 91 %), almost all participants who reported this theme stated that the robot agreed with the other more: *"We noticed that the robot was almost always disagreeing with my ideas and agreeing with his ideas"* (p.12, *negative feedback, unequal treatment*). In the *positive feedback* condition (22/24, 92 %), almost all participants who reported this theme stated that the robot agreed with me more: *"The robot really liked me and at the same time really hated her"* (p.5, *positive feedback, unequal treatment*).

Theme 2: Perception of the Other Participant. 114 out of 150 participants (76 %) discussed their perception of the other participant. Of the 114 participants, most participants (across all conditions)

mentioned positive aspects of the other participant. This positive perception was slightly more dominant in the *negative feedback and equal treatment* condition (32/35, 91 %) compared to the other conditions: *positive feedback and equal treatment* condition (25/29, 86 %), *positive feedback and unequal treatment* condition (20/23, 87 %), *negative feedback and unequal treatment* condition (23/27, 85 %). Participants referred to the atmosphere of the interaction: *"I realized that we have lots of agreements on many subjects, we are sharing the same vibe"* (p.10, *negative feedback, equal treatment*). They also mentioned the other participant's personality traits: *"I think he's a nice guy, personable and charismatic"* (p.2, *positive feedback, equal treatment*).

Theme 3: Perception of the Robot. 118 out of 150 participants (79 %) discussed their general perception of the robot. In *positive feedback and equal treatment* (13/30, 43 %), *negative feedback and equal treatment* (17/36, 47 %), and *positive feedback and unequal treatment* (12/26, 46 %) conditions, more participants described positive aspects of the robot than negative or neutral: *"It was like speaking to a person, not a robot. It felt like a human being"* (p.13, *positive feedback, equal treatment*). The *negative feedback and unequal treatment* condition was the only one where *negative* perception of the robot was more dominant (13/26, 50 %) when compared to positive and neutral: *"We were speaking to something that [likes] nothing"* (p.46, *negative feedback, unequal treatment*).

Theme 4: Perception of the Robot's Influence on Participant's Human-Human Interaction. 81 out of 150 participants (54 %) discussed the robot's influence on the human-human interaction. In both *equal treatment* conditions, *positive feedback* (10/19, 53 %) and *negative feedback* (10/19, 53 %), more than half of the participants who mentioned this theme described the robot having a positive impact: *"I find the robot to be very helpful, especially in situations that you feel awkward speaking with someone you don't know, the robot breaks the ice"* (p.123, *negative feedback, equal treatment*); *"I think that the robot shaped the mood of the conversation between us in a good way"* (p.125, *negative feedback, equal treatment*); *"I think that when he disagreed with us, it created some sort of connection between us"* (p.133, *negative feedback, equal treatment*). In the *unequal treatment* conditions, *positive feedback* (10/22, 45 %) and *negative feedback* (8/21, 38 %), less than half of the participants who mentioned this theme described a positive impact on their human-human interaction. In these conditions, participants described several negative aspects of the robot's impact: *"The robot created a strong sense of awkwardness"* (p.17, *positive feedback, unequal treatment*); *"The robot was a distraction, and he separated us"* (p.60, *positive feedback, equal treatment*).

4. Discussion

In decision-making tasks, robots have the potential to offer valuable data-driven feedback in the real world such as in education (Belpaeme et al., 2018; Hsieh et al., 2023; Rau et al., 2013; Sandoval et al., 2021), healthcare (Fasola and Mataric, 2012; Rea et al., 2021), business (Van Looy, 2020; Zhang, Tang, et al., 2023), and manufacturing (Gombolay et al., 2015; Kao and Liu, 2022). Due to their physical embodiment, they have the potential to influence people's behavior more strongly than screen-based and audio-based agents (Bainbridge et al., 2011; Deng et al., 2019; Howard and Borenstein, 2018; Wainer et al., 2007). However, little is known about how a robot's agreement or disagreement with people can influence the interpersonal closeness between human team members in two-person decision-making tasks. In this work, we found a significant influence of the robot's feedback on human-human relationships within the group. We further showed that this influence may depend on the robot's morphology and capabilities. Experiment 1 results indicate that a high anthropomorphism robot's *feedback valence* can decrease the sense of closeness between people when the robot gives *negative feedback* instead of *positive feedback*. Experiment 2 results show that a low anthropomorphism robot's *unequal treatment* as opposed to *equal treatment* and a low anthropomorphism robot's *positive feedback* as opposed to *negative feedback* can draw participants further apart.

4.1. Influence of the high anthropomorphism robot

We discovered that the high anthropomorphism robot's feedback valence significantly influenced interpersonal closeness between two people. When participants experienced *positive feedback* from the robot, they reported feeling closer to the other participant compared to those who experienced *negative feedback* from the robot. We further discovered that when participants were treated equally by the robot, the robot's *negative feedback* led to lower liking between the participants. Hence, when the entire interaction experience and the robot's responses towards both participants were characterized by negativity, the participants did not like each other as much. These results support the Influence of Negative Affect where a prevalence of negative affect and hostile effect could hurt human-human relationships in group settings (Gawronski and Walther, 2008; Gottman, 2014; Gottman and Levenson, 1992; Jung, 2016). This implies that we must not overlook the potential consequences of a robot's negative feedback, as it can pose a possible threat to human-human interpersonal closeness in decision-making teams.

Although the high anthropomorphism robot's negative feedback has an adverse effect on human-human interpersonal closeness, most participants did not mention it during their interviews and described the robot's influence as *neutral* (neither positive nor negative). This finding is important to highlight because robots that are incorporated into decision-making teams with multiple people may have an *implicit* negative influence on human-human relationships within the group without people in the group noticing it. Surprisingly, we found that a small number of participants in the *negative feedback and equal treatment* condition reported a *positive* influence of the robot on their relationship with the other participant, for example, stating that "*we were both laughing every time it disagreed with us, so that might have helped us connect a little bit more*" (p.45, *negative feedback, equal treatment)*. Future studies can investigate under which circumstances participants would consider the robot as their common enemy and leverage the robot's equal treatment to compensate for its negativity. Still, participant interviews demonstrate that a vast majority of the participants were unaware that the robot's negative feedback was harming their perceptions of closeness with the other participant, underscoring the importance of work like ours that aims to uncover the potential negative effects of robots and artificial agents on human-human interactions.

Our results also revealed that participants had a more favorable opinion of the robot when it gave *positive feedback* as opposed to *negative feedback*. When participants were asked about their perception of the robot, we found significantly more participants thought the robot was warmer and more competent when the robot gave *positive feedback* as opposed to *negative feedback*. These findings show that people expect robots in two-person decision-making tasks to follow social norms by being positive and polite to their users. Future designers of robots that are capable of giving opinions will need to carefully consider how to introduce the robot's constructive feedback to help people make decisions.

It is worth noting that the robot's positive and negative feedback towards the participants did not significantly affect the participants' rating of the robot's discomfort. Interestingly, the robot's treatment made participants rate the robot as significantly more discomforting when the robot treated them *unequally* as opposed to *equally*. This observation suggests that Heider's Balance Theory (Heider, 2013) may also apply in the two-person decision-making with the high anthropomorphism robot where the robot's imbalanced treatment is not preferred in a triadic relationship. However, the effect from Heider's Balance Theory is only notable in ratings of the robot's discomfort. The influence of negative affect from the high anthropomorphism robot is prominent in affecting ratings of the robot's warmth and competence as well as the interpersonal closeness between the two people in the two-person decision-making context.

4.2. Influence of the low anthropomorphism robot

We found that the low anthropomorphism robot's treatment (equal or unequal) of two people had a main effect on influencing human-human interpersonal closeness between the two participants. Participants receiving *equal treatment* from the robot felt closer with the other participant compared to those receiving *unequal treatment* from the robot. This finding is further supported by the interview results where more participants thought the robot had a *negative* effect on their interpersonal closeness when the robot gave *unequal treatment* as opposed to *equal treatment*. These results support Heider's Balance Theory (Heider, 2013), where people have more positive impressions and greater interpersonal closeness with each other when they agree with each other or when two of them are equally rejected by the third party (i.e., a common enemy). Specifically, when the robot treated both participants equally positively, participants felt more connected with one another because the robot equally agreed with their answers as opposed to breaking the balance by only agreeing with one participant's answer during the two-person decision-making task. When the robot treated both participants equally negatively, they also felt much closer to each other because they faced a common enemy, which was the robot that equally rejected both of their answers during the two-person decision-making task. Inferring from these results, we cannot overlook a robot's unequal treatment of people in a decision-making team because it may break the team's balance and detrimentally hurt interpersonal closeness.

Counter-intuitively, our results also indicated that when participants experienced *negative feedback* from the robot, they reported a closer relationship with one another than those who experienced *positive feedback* from the robot. This finding suggests that negative feedback from the simple low anthropomorphism robot encouraged people to seek closeness and enhance their relationship with the other participant. Unlike the influence of the high anthropomorphism robot in Experiment 1, the negativity presented by the low anthropomorphism robot did not lead to adverse effects on human-human interaction. Participants' tendency to seek closeness when experiencing *negative feedback* may represent their attempt to regulate the negative impact of the robot's feedback by increasing a sense of belonging with the other participant. This idea is supported by ostracism studies indicating that rejection by robots can increase people's need to belong (Erel et al., 2021) and encourage their behaviors that foster more social acceptance (Erel et al., 2022).

4.3. Lack of difference in the balance ratio from the decision-making outcomes

In both Experiment 1 and Experiment 2, we did not see the feedback from either of the two robots lead to differences in the balance ratio of how equally likely each participant's initial viewpoints got incorporated into the finalized joint answers. We postulate that since the robots did not give detailed reasoning on why they agreed or disagreed with a participant's answer (i.e., the high anthropomorphism robot only summarized what a participant had already said with pre-scripted agreeing and disagreeing statements, and the low anthropomorphism robot only gestured nodding and shaking), the participants may not have given as much consideration on the robot's feedback when making a joint answer, thus leading to a more balanced ratio across all types of pairs (i.e., Positive-Positive, Positive-Negative, Negative-Negative). This balanced ratio may also be explained by participants' tendency to conform to social norms by being polite and respectful to each other and trying to go with the other participant's answer or consider incorporating the other participant's answer into the final decision. While we did not see any influence of the robot's feedback on the likelihood of each participant's viewpoints being incorporated into the final joint answer, future studies can analyze decision-making outcomes when the robot gives more detailed reasoning for agreement or disagreement in a longer-term study.

4.4. Differences in responses to the high anthropomorphism and low anthropomorphism robots

We observed a striking difference in the effects of the high anthropomorphism robot (Experiment 1) and low anthropomorphism robot (Experiment 2) on participants' interpersonal closeness with the other participant. For the high anthropomorphism robot (Experiment 1), positive feedback as opposed to negative feedback resulted in greater interpersonal closeness. However, for the low anthropomorphism robot (Experiment 2), equal treatment as opposed to unequal treatment and negative feedback as opposed to positive feedback resulted in greater interpersonal closeness respectively. We postulate that the difference in the effects of these two robots is likely due to the higher intensity of the high anthropomorphism robot's negativity. While it is possible that Heider's Balance Theory (Heider, 2013) and the Influence of Negative Affect (Gawronski and Walther, 2008; Gottman, 2014; Gottman and Levenson, 1992; Jung, 2016) have both been at play, the intensity of the high anthropomorphism robot's negativity may have been so strong as to drown out any effect resulting from equal or unequal treatment. Additionally, the low anthropomorphism robot's negative feedback may have been so weak that its corrosive nature was not felt by participants as much as the imbalance created by the robot's unequal feedback. We focus on three factors we posit may have led to an increased intensity of the high anthropomorphism robot's negativity: the stronger negativity presented by the high anthropomorphism robot's speech, the humanness of the high anthropomorphism robot, and the distinct method that each robot used to deliver feedback.

While both the high and low anthropomorphism robots used head gestures to indicate agreement or disagreement, only the high anthropomorphism robot used verbal language to also convey its feedback. This negative feedback conveyed through speech may have made the high anthropomorphism robot's feedback seem much more intense compared to the simple negative gestures presented by the low anthropomorphism robot. For example, during the last question, the high anthropomorphism robot would shake its head and say "I am fundamentally opposed with your viewpoint," whereas the low anthropomorphism robot would only perform seven left-right shakes. The integration of GPT 3.5 into the high anthropomorphism robot's speech may have also captured more attention from the participants since the robot's speech was more personalized to their responses. This combination of the verbal statements of disagreement and personalized speech of the robot to the participants' responses could have led to the perception of the high anthropomorphism robot's feedback as more negative than that of the low anthropomorphism robot, leading to an overall negative atmosphere that resulted in participants feeling less close to one another.

The humanness of the high anthropomorphism robot, including its physical appearance and pre-assumed human-like capabilities, may have made the participants attribute more mind and perceived agency to the high anthropomorphism robot (Gray et al., 2007; Marchesi et al., 2022; Nijssen et al., 2019) and thus take its negative feedback more seriously compared to the low anthropomorphism robot. One participant who interacted with the high anthropomorphism robot confirmed the robot's humanness by mentioning: "I thought [the robot] was gonna be pre-programmed, but I guess it's just like a human, pretty much like a therapist" (p.18, *positive feedback, equal treatment*). Due to a lack of human-like presence from the low anthropomorphism robot, participants could more easily regulate its criticism by simply disregarding its negative attitude towards them and their ideas. One participant who experienced the low anthropomorphism robot's feedback mentioned: "At the end of the day it's a robot, it's not a human being. I think that if a robot disagrees with you it means a bit less because we don't actually know what it's doing. If humans disagree with you, you know that they're disagreeing with you, they make it very clear" (p.31, *negative*

feedback, unequal treatment). Another participant who interacted with the low anthropomorphism robot pointed out: "Well, it doesn't make you feel good when someone's shaking their head the entire time you're giving an answer. Yeah, so probably negative, but like, it's a robot, so I don't feel very strongly if that makes sense" (p.75, *negative feedback, unequal treatment*). As these quotations illustrate, the human-likeness of the robot may have impacted how much participants allowed the feedback of the robot to influence their interaction and perception with the other participant, leading to a stronger corrosive effect of the robot's negative feedback with the high anthropomorphism robot as opposed to the low anthropomorphism robot.

In addition, the difference in the way that the robot's feedback was provided between both experiments may have allowed participants to regulate their emotions differently and interpret the robot's feedback intensity distinctly. The high anthropomorphism robot gave feedback after participants finished delivering their answers and the low anthropomorphism robot gave feedback while the participants were giving their answers. Participants who interacted with the low anthropomorphism robot could have had a chance to justify their answers and regulate their emotions while the low anthropomorphism robot was giving its gestures. This discrepancy may have weakened the impact of the low anthropomorphism robot's negative feedback and thus lessened its impact on participants' experience, resulting in a different impact on their interpersonal closeness with the other participant. Despite the fact that we saw the Influence of Negative Affect endure with the high anthropomorphism robot (Experiment 1) and Heider's Balance Theory hold with the low anthropomorphism robot (Experiment 2), we believe that both psychological theories are likely always at play in decision-making groups and that the nature of the robot's feedback may make one more influential than the other.

4.5. Broader implications

Robots are becoming increasingly incorporated into our daily lives and demonstrate great potential in assisting people in team decision-making tasks (Booth et al., 2017; Erel et al., 2024; Fasola and Mataric, 2012; Hoffman, Forlizzi, et al., 2015; Hsieh et al., 2023; Mizumaru et al., 2019; Rea et al., 2021; Rosenthal and Veloso, 2012; Sandoval et al., 2021; Sebo et al., 2019). While incorporating robots into decision-making contexts may provide benefits to the group (Fasola and Mataric, 2012; Rea et al., 2021; Zhang, Tang, et al., 2023), our results show that robots may also negatively influence the relationship between the people in the group when they provide negative feedback (supporting the Influence of Negative Affect (Gawronski and Walther, 2008; Gottman, 2014; Gottman and Levenson, 1992; Jung, 2016)) or give imbalanced feedback (backed by Heider's Balance Theory (Heider, 2013)). Overlooking the intensity of the robot's negative feedback and the equality of the treatment given to the different group members may impair the interpersonal closeness between them and, as a result, change the nature of the group dynamics. Robots with negative feedback and unequal treatment may at some point also negatively impact the decisions taken by the group, since impairing group cohesion is known to eventually interfere with the quality of the group's decisions (Edmondson, 1999; Jones and George, 1998; Shore et al., 2011). This indicates that without accounting for the robot's feedback intensity and its balanced behavior toward the people in the group, the robot's impact may interfere with the decision-making process. We suggest that robot designers and practitioners should account for both aspects (feedback intensity and balanced treatment) when integrating robots into decision-making processes. For example, when designing robotic behavior for a robot that would facilitate brainstorming sessions in companies or medical teams, constructive negative comments from robots could be crucial to encourage the team to generate more innovative ideas or help prevent incorrect decisions. However, the robot designers need to be meticulous in employing negative speech from the robot and try to be mindful of a balanced robot treatment of all the individuals in the group setting.

Our findings also suggest that the negative effects of robots on two-person decision-making demonstrated in this work should be considered with robots of different levels of anthropomorphism. While the robots' anthropomorphism led to different effects, in both cases, both robots were able to influence the interpersonal closeness between the two participants. We highlight the need to design a balanced behavior toward people in the group, even when designing very simple machine-like robots that communicate only via minimal gestures. We further suggest that the intensity of negative feedback provided by a high anthropomorphism robot that can communicate using speech should be carefully considered and controlled. Given the rise of large language models in recent years, robots will soon be equipped with non-scripted verbal language to help people make decisions. Our research calls for careful consideration and control over the precise behaviors robots use to provide feedback.

5. Limitations

Even though we made every effort to ensure our two experiments were as similar as possible in experimental design while differing only in the level of robot anthropomorphism, there were also other differences between the two experiments to highlight as limitations. We conducted Experiment 1 and Experiment 2 in two different countries with separate participant populations. While both countries are considered to have a Western culture, participants in Experiment 1 did not include many international participants, whereas 40 % of the participants in Experiment 2 were international students from the university, representing a more diverse range of cultures. Additionally, Experiment 1 was conducted on a population that ranged more in age ($M = 33.06$, $SD = 12.42$) and was sourced from a prominent behavioral science museum in a major city, thus attracting a pool of participants with more diverse backgrounds. Experiment 2 was conducted primarily on university students on campus with less variability in age ($M = 24.23$, $SD = 7.04$) and background. While these differences in our participant populations for Experiment 1 and Experiment 2 may have contributed to the variance in the results, we note that we have a large sample size (Experiment 1: N = 172; Experiment 2: N = 150) and ground our research in relevant psychological theories (Gawronski and Walther, 2008; Gottman, 2014; Gottman and Levenson, 1992; Heider, 2013; Jung, 2016) that were validated across different cultures. We would also like to point out that we did not ask participants in Experiment 1 to fill out pre-study questionnaires such as the Attitude Towards Robots questionnaire (Nomura et al., 2006) a few days in advance, similar to the procedure in Experiment 2, because the participants in Experiment 1 were all museum visitors who sought to join the study on the same day of their museum visit. We chose not to ask them to fill out pre-study questionnaires right before the study to avoid priming effects. At the cost of not matching the procedure of filling out pre-study questionnaires in Experiment 2, we chose to conduct the study with a more diverse participant pool.

6. Conclusion

In this work, we launched two parallel experiments to examine the impact of the robot's feedback on interpersonal closeness between two participants in a two-person decision-making task. Experiment 1 used a high anthropomorphism robot and showed its feedback valence could enhance interpersonal closeness between two people when the robot gave positive feedback as opposed to negative feedback. Experiment 2 used a low anthropomorphism robot and showed that its treatment of two people could lead to greater interpersonal closeness between two people when the robot provided equal feedback instead of unequal feedback. The low anthropomorphism robot's feedback valence also increased interpersonal closeness between two people when the robot gave negative feedback as opposed to positive feedback. Future robot designers and practitioners should consider both the robot's feedback

and treatment when introducing them to tasks involving two-person decision-making.

CRediT authorship contribution statement

Ting-Han Lin: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Yuval Rubin Kopelman:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Madeline Busse:** Writing – review & editing, Writing – original draft, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Sarah Sebo:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Hadas Erel:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

1. Acknowledgments

All authors of this manuscript were involved in the phases of conceptualization, methodology design, user study deployment, data analysis, and writing. We would like to thank Esha Mujumdar, Noa Kirtchuk, Mai Efrat, Hila Zohar, Andrey Grishko, Nevo Heimann Saadon, and Eden Lulu for their research assistance. We would also like to thank the staff at Mindworks in Chicago for assisting us in recruiting participants in Experiment 1. This work was supported by research funds at the University of Chicago, the National Science Foundation (Award #2312352), and Reichman University.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chb.2025.108807.

Data availability

We cited and linked our data in this paper.

References

Abdi, E., Tojib, D., Seong, A. K., Pamarthi, Y., & Millington-Palmer, G. (2022). A study on the influence of service robots' level of anthropomorphism on the willingness of users to follow their recommendations. *Scientific Reports, 12*, Article 15266.

Anderson-Bashan, L., Megidish, B., Erel, H., Wald, I., Hoffman, G., Zuckerman, O., & Grishko, A. (2018). The greeting machine: An abstract robotic object for opening encounters. In *2018 27th IEEE international symposium on robot and human interactive communication (RO-MAN)* (pp. 595–602). IEEE.

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology, 63*, 596.

Aron, A., Melinat, E., Aron, E. N., Vallone, R. D., & Bator, R. J. (1997). The experimental generation of interpersonal closeness: A procedure and some preliminary findings. *Personality and Social Psychology Bulletin, 23*, 363–377.

Bainbridge, W. A., Hart, J. W., Kim, E. S., & Scassellati, B. (2011). The benefits of interactions with physically present robots over video-displayed agents. *International Journal of Social Robotics, 3*, 41–52.

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., & Tanaka, F. (2018). Social robots for education: A review. *Science Robotics, 3*, Article eaat5954.

Birnbaum, G. E., Mizrahi, M., Hoffman, G., Reis, H. T., Finkel, E. J., & Sass, O. (2016). What robots can teach us about intimacy: The reassuring effects of robot responsiveness to human disclosure. *Computers in Human Behavior, 63*, 416–423.

Booth, S., Tompkin, J., Pfister, H., Waldo, J., Gajos, K., & Nagpal, R. (2017). Piggybacking robots: Human-robot overtrust in university dormitory security. In *Proceedings of the 2017 ACM/IEEE international conference on human-robot interaction* (pp. 426–434).

Bosson, J. K., Johnson, A. B., Niederhoffer, K., & Swann Jr, W. B. (2006). Interpersonal chemistry through negativity: Bonding by sharing negative attitudes about others. *Personal Relationships, 13*, 135–150.

Boyatzis, R. (1998). *Transforming qualitative information: Thematic analysis and code development*. Sage.

Burgoon, J. K., & Hale, J. L. (1984). The fundamental topoi of relational communication. *Communication Monographs, 51*, 193–214.

Carpinella, C. M., Wyman, A. B., Perez, M. A., & Stroessner, S. J. (2017). The robotic social attributes scale (rosas) development and validation. In *Proceedings of the 2017 ACM/IEEE international conference on human-robot interaction* (pp. 254–262).

Claure, H., Chen, Y., Modi, J., Jung, M., & Nikolaidis, S. (2020). Multi-armed bandits with fairness constraints for distributing resources to human teammates. In *Proceedings of the 2020 ACM/IEEE international conference on human-robot interaction* (pp. 299–308).

Claure, H., Kim, S., Kizilcec, R. F., & Jung, M. (2023). The social consequences of machine allocation behavior: Fairness, interpersonal perceptions and performance. *Computers in Human Behavior, 146*, Article 107628.

Coglianese, C., & Lehr, D. (2016). Regulating by robot: Administrative decision making in the machine-learning era. *Geological Journal, 105*, 1147.

Deng, E., Mutlu, B., & Mataric, M. J. (2019). Embodiment in socially interactive robots. *Foundations and Trends® in Robotics, 7*, 251–356.

Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems, 42*, 177–190.

Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly, 44*, 350–383.

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review, 114*, 864.

Erel, H., Carsenti, E., & Zuckerman, O. (2022). A carryover effect in hri: Beyond direct social effects in human-robot interaction. In *2022 17th ACM/IEEE international conference on human-robot interaction (HRI)* (pp. 342–352). IEEE.

Erel, H., Cohen, Y., Shafrir, K., Levy, S. D., Vidra, I. D., Shem Tov, T., & Zuckerman, O. (2021). Excluded by robots: Can robot-robot-human interaction lead to ostracism?. In *Proceedings of the 2021 ACM/IEEE international conference on human-robot interaction* (pp. 312–321).

Erel, H., Shem Tov, T., Kessler, Y., & Zuckerman, O. (2019). Robots are always social: Robotic movements are automatically interpreted as social cues. In *Extended abstracts of the 2019 CHI conference on human factors in computing systems* (pp. 1–6).

Erel, H., Vázquez, M., Sebo, S., Salomons, N., Gillet, S., & Scassellati, B. (2024). Rosi: A model for predicting robot social influence. *ACM Transactions on Human-Robot Interaction, 13*(2), 1–22.

Fasola, J., & Mataric, M. J. (2012). Using socially assistive human–robot interaction to motivate physical exercise for older adults. *Proceedings of the IEEE, 100*, 2512–2526.

Galletta, A. (2013). *Mastering the semi-structured interview and beyond: From research design to analysis and publication, ume 18*. NYU press.

Gawronski, B., & Walther, E. (2008). The tar effect: When the ones who dislike become the ones who are disliked. *Personality and Social Psychology Bulletin, 34*, 1276–1289.

Gibbs, G. (2008). *Analysing qualitative data (qualitative research kit)*. Retrieved from.

Gillet, S., Cumbal, R., Pereira, A., Lopes, J., Engwall, O., & Leite, I. (2021). Robot gaze can mediate participation imbalance in groups with different skill levels. In *Proceedings of the 2021 ACM/IEEE international conference on human-robot interaction* (pp. 303–311).

Gillet, S., Vázquez, M., Andrist, S., Leite, I., & Sebo, S. (2024). Interaction-shaping robotics: Robots that influence interactions between other agents. *ACM Transactions on Human-Robot Interaction, 13*, 1–23.

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *The Academy of Management Annals, 14*, 627–660.

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

Gottman, J. M. (2014). *What predicts divorce?: The relationship between marital processes and marital outcomes*. Psychology Press.

Gottman, J. M., & Levenson, R. W. (1992). Marital processes predictive of later dissolution: Behavior, physiology, and health. *Journal of Personality and Social Psychology, 63*, 221.

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science, 315*, 619.

Guzman, A. L. (2016). *The messages of mute machines: Human-machine communication with industrial technologies, 1 5*. communication+.

Heider, F. (2013). *The psychology of interpersonal relations*. Psychology Press.

Hoffman, G., Forlizzi, J., Ayal, S., Steinfeld, A., Antanitis, J., Hochman, G., Hochendoner, E., & Finkenaur, J. (2015). Robot presence and human honesty: Experimental evidence. In *Proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction* (pp. 181–188).

Hoffman, G., & Ju, W. (2014). Designing robots with movement in mind. *Journal of Human-Robot Interaction, 3*, 91–122.

Hoffman, G., Zuckerman, O., Hirschberger, G., Luria, M., & Shani Sherman, T. (2015). Design and evaluation of a peripheral robotic conversation companion. In *Proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction* (pp. 3–10).

Howard, A., & Borenstein, J. (2018). The ugly truth about ourselves and our robot creations: The problem of bias and social inequity. *Science and Engineering Ethics, 24*, 1521–1536.
Hsieh, T. Y., Chaudhury, B., & Cross, E. S. (2023). Human–robot cooperation in economic games: People show strong reciprocity but conditional prosociality toward robots. *International Journal of Social Robotics, 15*, 791–805.
Huang, H., & Liu, S. Q. (2022). Are consumers more attracted to restaurants featuring humanoid or non-humanoid service robots? *International Journal of Hospitality Management, 107*, Article 103310.
Huang, H., Rau, P. L. P., & Ma, L. (2021). Will you listen to a robot? Effects of robot ability, task complexity, and risk on human decision-making. *Advanced Robotics, 35*, 1156–1166.
Iwasaki, M., Yamazaki, A., Yamazaki, K., Miyazaki, Y., Kawamura, T., & Nakanishi, H. (2024). Perceptive recommendation robot: Enhancing receptivity of product suggestions based on customers' nonverbal cues. *Biomimetics, 9*, 404.
Jacobs, M., Pradier, M. F., McCoy Jr, T. H., Perlis, R. H., Doshi-Velez, F., & Gajos, K. Z. (2021). How machine-learning recommendations influence clinician treatment selections: The example of antidepressant selection. *Translational Psychiatry, 11*, 108.
Jain, P., Coogan, S. C., Subramanian, S. G., Crowley, M., Taylor, S., & Flannigan, M. D. (2020). A review of machine learning applications in wildfire science and management. *Environmental Reviews, 28*, 478–505.
Jones, G. R., & George, J. M. (1998). The experience and evolution of trust: Implications for cooperation and teamwork. *Academy of Management Review, 23*, 531–546.
Ju, W., & Takayama, L. (2009). Approachability: How people interpret automatic door movement as gesture. *International Journal of Design, 3*.
Jung, M. F. (2016). Coupling interactions and performance: Predicting team performance from thin slices of conflict. *ACM Transactions on Computer-Human Interaction, 23*, 1–32.
Jung, M. F., DiFranzo, D., Shen, S., Stoll, B., Claure, H., & Lawrence, A. (2020). Robot-assisted tower construction—a method to study the impact of a robot's allocation behavior on interpersonal dynamics and collaboration in groups. *ACM Transactions on Human-Robot Interaction (THRI), 10*, 1–23.
Jung, M. F., DiFranzo, D., Stoll, B., Shen, S., Lawrence, A., & Claure, H. (2018). Robot-assisted tower construction—a resource distribution task to study human–robot collaboration and interaction with groups of people. *arXiv preprint arXiv:1812.09548*.
Jung, M. F., Martelaro, N., & Hinds, P. J. (2015). Using robots to moderate team conflict: The case of repairing violations. In *Proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction* (pp. 229–236).
Kao, C., & Liu, S. T. (2022). Group decision making in data envelopment analysis: A robot selection application. *European Journal of Operational Research, 297*, 592–599.
Lewis, M., Sycara, K., & Walker, P. (2018). The role of trust in human-robot interaction. *Foundations of trusted autonomy*, 135–159.
Liberman-Pincu, E., van Grondelle, E. D., & Oron-Gilad, T. (2022). Designing robots with the context in mind-one design does not fit all. In *International workshop on human-friendly robotics* (pp. 105–119). Springer.
Ludwig, V. U., Berry, B., Cai, J. Y., Chen, N. M., Crone, D. L., & Platt, M. L. (2022). The impact of disclosing emotions on ratings of interpersonal closeness, warmth, competence, and leadership ability. *Frontiers in Psychology, 13*, Article 989826.
Manor, A., Megidish, B., Todress, E., Mikulincer, M., & Erel, H. (2022). A nonhumanoid robotic object for providing a sense of security. In *2022 31st IEEE international conference on robot and human interactive communication (RO-MAN)* (pp. 1520–1527). IEEE.
Marchesi, S., Tommaso, D. D., Perez-Osorio, J., & Wykowska, A. (2022). Belief in sharing the same phenomenological experience increases the likelihood of adopting the intentional stance toward a humanoid robot. *Technology, Mind, and Behavior, 3*.
Masjutin, L., Laing, J. K., & Maier, G. W. (2022). Why do we follow robots? An experimental investigation of conformity with robot, human, and hybrid majorities. In *2022 17th ACM/IEEE international conference on human-robot interaction (HRI)* (pp. 139–146). IEEE.
Maxwell, G. M., Cook, M. W., & Burr, R. (1985). The encoding and decoding of liking from behavioral cues in both auditory and visual channels. *Journal of Nonverbal Behavior, 9*, 239–263.
Megidish, B. (2017). Butter robotics. URL: https://butter-robotics.web.app/.
Mizumaru, K., Satake, S., Kanda, T., & Ono, T. (2019). Stop doing it! approaching strategy for a robot to admonish pedestrians. In *2019 14th ACM/IEEE international conference on human-robot interaction (HRI)* (pp. 449–457). IEEE.
Nakanishi, H., Nakazawa, S., Ishida, T., Takanashi, K., & Isbister, K. (2003). Can software agents influence human relations? Balance theory in agent-mediated communities. In *Proceedings of the second international joint conference on autonomous agents and multiagent systems* (pp. 717–724).
Nijssen, S. R., Müller, B. C., Baaren, R. B.v., & Paulus, M. (2019). Saving the robot or the human? Robots who feel deserve moral care. *Social Cognition, 37*, 41.
Nomura, T., Kanda, T., & Suzuki, T. (2006). Experimental investigation into influence of negative attitudes toward robots on human–robot interaction. *AI & Society, 20*, 138–150.
Ososky, S., Schuster, D., Phillips, E., & Jentsch, F. G. (2013). Building appropriate trust in human-robot teams. In *AAAI spring symposium: Trust and autonomous systems* (pp. 60–65).
Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39*, 230–253.
Polakow, T., Laban, G., Teodorescu, A., Busemeyer, J. R., & Gordon, G. (2022). Social robot advisors: Effects of robot judgmental fallacies and context. *Intelligent Service Robotics, 15*, 593–609.
Rajagopal, N. K., Qureshi, N. I., Durga, S., Ramirez Asis, E. H., Huerta Soto, R. M., Gupta, S. K., & Deepak, S. (2022). Future of business culture: An artificial intelligence-driven digital framework for organization decision-making process. *Complexity, 2022*, Article 7796507.
Rau, P. L. P., Li, Y., & Liu, J. (2013). Effects of a social robot's autonomy and group orientation on human decision-making. *Advances in human-computer interaction 2013*, Article 263721.
Rea, D. J., Schneider, S., & Kanda, T. (2021). " is this all you can do? harder!" the effects of (im) polite robot encouragement on exercise effort. *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, 225–233.
Rosenthal, S., & Veloso, M. (2012). Mobile robot planning to seek help with spatially-situated tasks. In *Proceedings of the AAAI conference on artificial intelligence* (pp. 2067–2073).
Sakamoto, D., & Ono, T. (2006). Sociality of robots: Do robots construct or collapse human relations?. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on human-robot interaction* (pp. 355–356).
Salomons, N., Sebo, S. S., Qin, M., & Scassellati, B. (2021). A minority of one against a majority of robots: Robots cause normative and informational conformity. *ACM Transactions on Human-Robot Interaction (THRI), 10*, 1–22.
Salomons, N., Van Der Linden, M., Strohkorb Sebo, S., & Scassellati, B. (2018). Humans conform to robots: Disambiguating trust, truth, and conformity. In *Proceedings of the 2018 acm/ieee international conference on human-robot interaction* (pp. 187–195).
Sandoval, E. B., Brandstatter, J., Yalcin, U., & Bartneck, C. (2021). Robot likeability and reciprocity in human robot interaction: Using ultimatum game to determinate reciprocal likeable robot strategies. *International Journal of Social Robotics, 13*, 851–862.
Schömbs, S., Pareek, S., Goncalves, J., & Johal, W. (2024). Robot-assisted decision-making: Unveiling the role of uncertainty visualisation and embodiment. In *Proceedings of the CHI conference on human factors in computing systems* (pp. 1–16).
Sebo, S. S., Krishnamurthi, P., & Scassellati, B. (2019). "i don't believe you": Investigating the effects of robot trust violation and repair. In *2019 14th ACM/IEEE international conference on human-robot interaction (HRI)* (pp. 57–65). IEEE.
Sebo, S., Stoll, B., Scassellati, B., & Jung, M. F. (2020). Robots in groups and teams: A literature review. *Proceedings of the ACM on Human-Computer Interaction, 4*, 1–36.
Shen, S., Slovak, P., & Jung, M. F. (2018). " stop. i see a conflict happening." a robot mediator for young children's interpersonal conflict resolution. In *Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction* (pp. 69–77).
Shinozawa, K., Naya, F., Yamato, J., & Kogure, K. (2005). Differences in effect of robot and screen agent recommendations on human decision-making. *International Journal of Human-Computer Studies, 62*, 267–279.
Shiomi, M., & Hagita, N. (2016). Do synchronized multiple robots exert peer pressure?. In *Proceedings of the fourth international conference on human agent interaction* (pp. 27–33).
Shiomi, M., & Hagita, N. (2019). Do the number of robots and the participant's gender influence conformity effect from multiple robots? *Advanced Robotics, 33*, 756–763.
Shore, L. M., Randel, A. E., Chung, B. G., Dean, M. A., Holcombe Ehrhart, K., & Singh, G. (2011). Inclusion and diversity in work groups: A review and model for future research. *Journal of Management, 37*, 1262–1289.
Stroessner, S. J., & Benitez, J. (2019). The social perception of humanoid and non-humanoid robots: Effects of gendered and machinelike features. *International Journal of Social Robotics, 11*, 305–315.
Tennent, H., Shen, S., & Jung, M. (2019). Micbot: A peripheral robotic object to shape conversational dynamics and team performance. *14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 133–142.
Van Looy, A. (2020). Adding intelligent robots to business processes: A dilemma analysis of employees' attitudes. In *Business process management: 18th international conference, BPM 2020, Seville, Spain, September 13–18, 2020* (pp. 435–452). Springer. Proceedings 18.
Wainer, J., Feil-Seifer, D. J., Shell, D. A., & Mataric, M. J. (2007). Embodiment and human-robot interaction: A task-based perspective. In *RO-MAN 2007-The 16th IEEE international symposium on robot and human interactive communication* (pp. 872–877). IEEE.
Waytz, A., Cacioppo, J., & Epley, N. (2010). Who sees human? The stability and importance of individual differences in anthropomorphism. *Perspectives on Psychological Science, 5*, 219–232.
Zhang, A. W., Lin, T. H., Zhao, X., & Sebo, S. (2023). Ice-breaking technology: Robots and computers can foster meaningful connections between strangers through in-person conversations. In *Proceedings of the 2023 CHI conference on human factors in computing systems* (pp. 1–14).
Zhang, S., Tang, G., Li, X., & Ren, A. (2023). The effects of appearance personification of service robots on customer decision-making in the product recommendation context. *Industrial Management & Data Systems, 123*, 578–595.
Zuckerman, O., Walker, D., Grishko, A., Moran, T., Levy, C., Lisak, B., Wald, I. Y., & Erel, H. (2020). Companionship is not a function: The effect of a novel robotic object on healthy older adults' feelings of" being-seen". In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1–14).