Can you pass that tool?: Implications of Indirect Speech in Physical Human-Robot Collaboration

Yan Zhang
School of Computing and Information Systems
University of Melbourne
Melbourne, VIC, Australia
yan.zhang.1@unimelb.edu.au

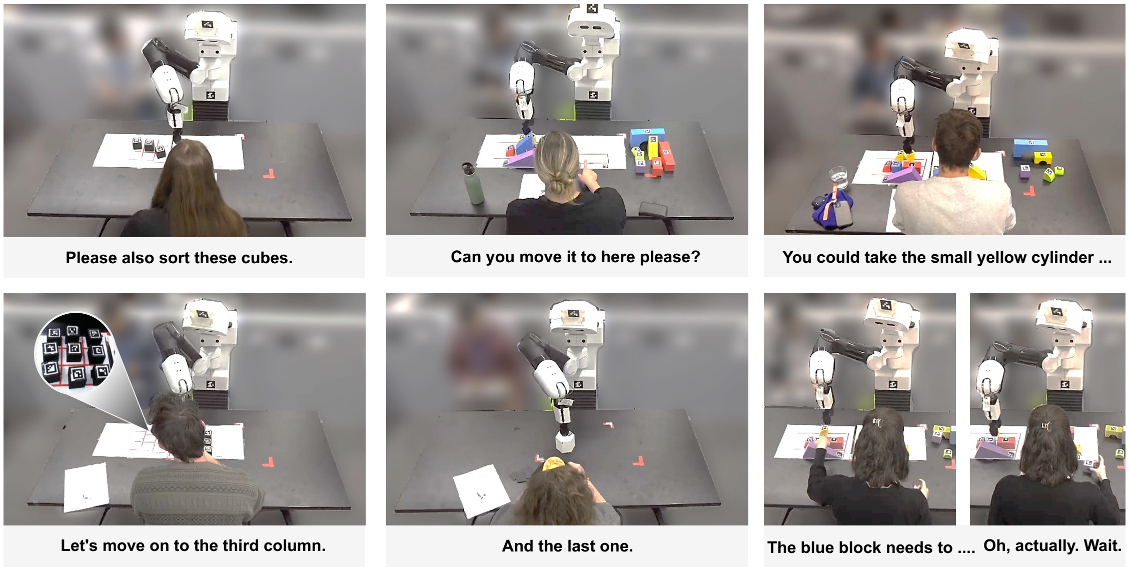
Tharaka Sachintha Ratnayake
University of Melbourne
Melbourne, VIC, Australia
tsratnayakem@student.unimelb.edu.au

Cherie Sew
School of Computing and Information Systems
University of Melbourne
Melbourne, VIC, Australia
csew@student.unimelb.edu.au

Jarrod Knibbe
School of Electrical Engineering and Computer Science
The University of Queensland
Brisbane, QLD, Australia
j.knibbe@uq.edu.au

Jorge Goncalves
School of Computing and Information Systems
University of Melbourne
Melbourne, VIC, Australia
jorge.goncalves@unimelb.edu.au

Wafa Johal
School of Computing and Information Systems
University of Melbourne
Melbourne, VIC, Australia
wafa.johal@unimelb.edu.au



Please also sort these cubes.

Can you move it to here please?

You could take the small yellow cylinder ...

Let's move on to the third column.

And the last one.

The blue block needs to ....

Oh, actually. Wait.

Figure 1: This figure presents images from our experiment, featuring representative participant utterances to illustrate the types of requests used. The top left image depicts a direct request, while the rest of the images showcase various indirect requests. The interpretation and robot's responses are explained in section 3.

Abstract

Indirect speech acts (ISAs) are a natural pragmatic feature of human communication, allowing requests to be conveyed implicitly while maintaining subtlety and flexibility. Although advancements in speech recognition have enabled natural language interactions with robots through direct, explicit commands—providing clarity in communication—the rise of large language models presents the potential for robots to interpret ISAs. However, empirical evidence on the effects of ISAs on human-robot collaboration (HRC) remains limited. To address this, we conducted a Wizard-of-Oz study (N=36), engaging a participant and a robot in collaborative physical tasks. Our findings indicate that robots capable of understanding ISAs significantly improve human's perceived robot anthropomorphism, team performance, and trust. However, the effectiveness of ISAs is task- and context-dependent, thus requiring careful use. These results highlight the importance of appropriately integrating direct and indirect requests in HRC to enhance collaborative experiences and task performance.

CHI '25, Yokohama, Japan
© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3713780

CCS Concepts

• **Human-centered computing → Empirical studies in HCI**; **User studies**.

Keywords

Human-Robot Collaboration, Language Communication, Grounding, Lab Study

ACM Reference Format:
Yan Zhang, Tharaka Sachintha Ratnayake, Cherie Sew, Jarrod Knibbe, Jorge Goncalves, and Wafa Johal. 2025. Can you pass that tool?: Implications of Indirect Speech in Physical Human-Robot Collaboration . In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26-May 1, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3706598.3713780

1 Introduction

A spoken sentence is often not limited to its literal meaning. The question *"Can you pass the salt?"* implicitly requests an action, while literally questioning the listener's physical ability to handover the salt. Alternatively, *"This soup needs salt"* both asserts an opinion about the soup and, depending on the context and surrounding objects, may be requesting someone to pass the salt [24]. These are examples of indirect speech acts (ISAs), which Searle [82] defined as utterances where one speech act is performed indirectly by carrying out another, transforming direct intents into implicatures. They are complex, multi-faceted, and require shared context and interpretation [82]. They are also an optimized way to communicate that commonly occurs in collaboration settings where teammates build a shared understanding of the task [26]. Similar to human-human interaction, understanding ISA in human-robot interaction is crucial, since interactions are often based on language to achieve a certain task or goal.

The inherent naturalness, ease of production, and flexibility of indirect speech make it well-suited for effective human-robot collaboration (HRC) [52, 66]. Through this lens, the robot is envisioned as a social, intelligent collaborator, where politeness, social etiquette, and discussion become factors of shared tasks. In emerging social collaborative robotic (cobotic) scenarios, such as with personal assistants in healthcare and accessibility [19, 112], then, ISAs are seen as a suitable method of interaction [104].

In performance- and task-oriented settings, however, the appropriateness of ISAs is less obvious. If the cobot partner is performing critical tasks, there may be little room for interpretation or lack of clarity. Direct speech – *"pass the salt"* – is clear and timely. This is akin to more traditional interactions, where robots were clearly subservient and commands needed to be learned and delivered correctly. However, this learning creates barriers to natural and intuitive interaction, and the influence on user experience lacks evidence from comparative user studies.

Even though the recent advances in large language models (LLMs) enhance the potential for natural speech interactions with robots in physical tasks [50, 53, 89, 113], to date, much of the attention is still on direct, explicit commands. This often oversimplifies communication, stripping away the naturalness observed in genuine human collaboration [47]. The rate of LLMs' advances makes it likely that indirect speech will be supported through these models, before which, however, it remains essential to understand the role and impact of ISAs in human-robot collaboration.

Previous work has shown that humans tend to use ISAs when interacting with robots at frequencies similar to those used with other humans [7, 48]. This highlights the need to develop human-centred verbal communication interfaces for cobots that can accommodate the naturalistic and varied ways in which people express themselves. However, despite the indispensability of ISAs in collaborative communication, there remains a gap in empirical evidence regarding the impact of ISAs on human-robot collaboration, especially in tabletop manipulation tasks.

To address this gap, we conducted a study with 36 participants, comparing two speech modes of a real robot in a laboratory setting: one capable of understanding ISAs and another without this capability, on three collaborative tasks. Given that natural language communication is a barrier preventing human-robot teams from outperforming human-human teams [80], we theorise that the use of ISAs can contribute to the effectiveness and naturalness of communication, thereby improving perceived team performance and user experience. Specifically, to assess the impact of ISAs on collaboration and communication, we evaluated four key metrics commonly used in HRC. *Team fluency* reflects seamless coordination, which is critical for user satisfaction and acceptance of cobots [31, 40]. *Goal alignment* measures the success and efficiency of collaboration [74, 75]. *Trust* is essential for preventing misuse or disuse of the robot, ultimately enhancing collaboration effectiveness [1, 109]. Moreover, enabling the robot's ability to understand ISAs could serve as a means to induce *anthropomorphism*, thereby improving collaborative engagement by fostering a sense of partnership and enhancing the collaborative experience [69, 110].

In summary, our research addresses the following questions:

- **RQ1** How does a robot's capability to understand indirect speech acts influence the perceived *team's performance*?
  - **RQ1.1** How does a robot's capability to understand indirect speech acts influence the *fluency* of human-robot teamwork?
  - **RQ1.2** How does a robot's capability to understand indirect speech acts influence the establishment of *goal alignment* among the human-robot team?
- **RQ2** How does a robot's capability to understand indirect speech acts influence a human teammate's *trust* in the robot's performance?
- **RQ3** How does a robot's capability to understand indirect speech acts influence a human teammate's perception of the robot's *anthropomorphism*?

Our findings show that while ISAs are beneficial in human-robot collaboration, their effectiveness can vary depending on the context. The quantitative results show the robot's ability to comprehend ISAs significantly enhances participants' perceived team performance, trust, and anthropomorphism. The use of ISAs fosters a deeper cognitive engagement, making the robot appear more as a collaborative partner rather than a mere tool. However, qualitative results suggest that the usage of ISA can be task- and context-dependent in human-robot collaboration, with inappropriate use potentially leading to negative impacts on trust and user perception. These insights highlight the inherent limitations of relying solely on direct command-based interactions, which lack the subtlety required for establishing shared understanding and the sense of teaming. They also emphasise the importance of using indirect requests in a contextually adaptive and appropriate manner. Therefore, the careful integration of direct and indirect verbal communication emerges as a critical factor in optimising the performance and overall experience of human-robot collaboration. We advocate for the human-computer interaction (HCI) and human-robot interaction (HRI) community to develop human-centred LLMs for collaborative robots, recognising the critical role of ISAs in achieving this goal.

2 Related Work

2.1 Speaking to Embodied Agents

Voice assistants (VA), like Siri and Alexa, can have a noticeable influence on user behaviour, as these voice command interfaces are increasingly integrated into daily interactions through devices like phones, computers, and cars. [4, 99]. The reach of VAs extends beyond simple task execution, influencing users' linguistic habits and potentially shaping social norms surrounding technology use [60, 104]. Early research primarily addressed the technical challenges associated with speech detection and dialogue systems, focusing on improving the accuracy and efficiency of voice recognition technologies [27]. As VAs became more commercialised, researchers observed that people adapt their language when interacting with VAs, using direct commands, simplified sentences, and keywords to mitigate the risk of misinterpretation [43, 61]. This adaptation reflects the users' low expectations of language processing and voice interfaces, as well as the inherent limitation of VAs' ability to comprehend and execute complex commands accurately. However, with the advancements in natural language processing, there has been a shift away from command-based paradigms towards more nuanced and complex verbal interactions due to its increased ability to infer intention and understand context [64, 94], allowing for more natural and fluid dialogues between users and machines [2, 107].

Voice command interfaces have been implemented in embodied agents, such as social and collaborative robots, offering significant advantages in making these systems more human-like assistants capable of supporting real-world tasks. The advantages of incorporating voice interfaces into robots are evident, particularly in scenarios where natural and intuitive communication is essential. Besides, voice interfaces make AI and digital information more accessible to specific populations, such as children and the elderly, who may otherwise struggle with traditional interaction methods [72, 87]. To make the voice interface more capable and better accommodate human activities, recent studies have increasingly focused on elements such as vocal fillers [67], voice-matching [30], and social norms, particularly language politeness [104]. Among these, most elements contribute to making interactions feel more natural and human-like. Anthropomorphism remains one of the most extensively studied characteristics in human-agent verbal interaction [81].

A considerable amount of existing research has explored the potential impact of robots' voices on human perception of human-likeness, trust, and capability. For instance, studies have shown that a human-like voice can increase trust in the robot [106], with this effect being more noticeable when the robot's voice matches the gender of the participant [34]. Another study found that participants issued more commands to robots with artificial voices compared to those with human-like speech, suggesting that a less human-like voice may lead users to perceive the robot as a less capable machine rather than as a competent human [88]. While the anthropomorphism of robots' speech can enhance user experience, it also increases the risk of participants overestimating the robot's intelligence and abilities [20]. Other factors like politeness, humor, and directness also shape a robot's perceived anthropomorphism [33]. Robots using indirect language in social interactions often seem more human-like [79].

2.2 Verbal Communication during Human-Robot Collaboration

Verbal communication offers distinct advantages due to its naturalness and efficiency [52]. Previous research has demonstrated that humans communicating task-related information to robots can enhance the robot's understanding of goals and intentions, thereby improving overall performance [14]. Additionally, studies have shown that robots equipped with communication abilities and verbal feedback can improve team performance by reducing task completion times and being perceived as better teammates [92]. Furthermore, explicitly incorporating context into communication enhances clarity, reduces ambiguity, and improves mutual understanding [58, 95].

In natural language processing for human-robot collaboration, several methods exist to parse commands from explicit utterances. The most direct approach involves extracting semantic features and mapping them to predefined robot controllers [96]. However, research has shown that participants often provide instructions at varying levels of abstraction [5]. To interpret more abstract commands lacking specific keywords, association models are used to combine literal linguistic features and extract semantic meaning, typically relying on probability-based methods [51, 57, 59]. Additionally, to generalize across new tasks and enable contextual understanding, researchers are exploring the usage of large language models for controlling robots in physical tasks [50, 53, 89, 113]. While LLMs have the potential to comprehend implicit verbal commands, most studies focus on explicit, direct commands, which provide clear instructions but do not capture the nuanced and indirect nature of human communication in real-world scenarios.

However, relying primarily on direct commands that explicitly convey human requests oversimplifies interactions. Indirect speech acts are a natural feature of human communication, contributing to enhancing robots' anthropomorphism, which has been shown to be an important factor in creating an ideal AI teammate [110]. ISAs also serve as an implicit and important means for humans to express their intentions. When the ISAs are misinterpreted during collaboration, the potential for long-term efficiency gains is compromised. Therefore, equipping robots with the ability to interpret ISAs enables them to respond more naturally and effectively, closely mimicking human-like communication patterns. This capability is particularly important in tasks that require high levels of coordination and mutual understanding, such as cooperative manipulation tasks [86].

2.3 Indirect Speech Acts in Human-Robot Interaction

Research has shown that humans tend to use indirect verbal requests when interacting with robots at frequencies similar to those used in human-human interactions, demonstrating the necessity of enabling ISAs in HRI [7, 48]. Several studies have focused on providing robots with the ability to interpret indirect requests. For example, Briggs and Scheutz [15] created a hybrid system to comprehend indirect requests and provide appropriate responses. Another studies [102, 103] introduced a probabilistic algorithm for robots to learn sociocultural norms, infer intentions from human utterances, and generate clarification requests.

Moreover, the impact of ISAs on human-robot interaction has been examined from various perspectives. Research shows that robots employing ISAs are perceived as more likeable [93, 98], trustworthy [79], and willing to help [91]. Conversely, a robot's inability to understand conventionalised ISAs (e.g., *"Can you..?"*, *"I need you to.."*) during social interactions has been found to negatively affect its performance and human perception [105]. Even when participants are aware that robots may not fully comprehend ISAs, they tend to continue using them, which can impact the interaction fluency [16]. While current research largely focuses on social interactions, there is still a limited exploration of ISAs in the context of physical collaboration, where conversations tend to be more collaborative, continuous, and shaped by physical context, rather than purely by politeness and social norms.

Although ISAs enable robots to engage in more nuanced and contextually rich social interactions, there is still a gap in the empirical evaluation of ISAs' impact on perceived task performance and user experience in HRC across physical collaborative tasks. In this study, we investigate the effects of a robot's ability to understand ISAs on team fluency, goal alignment, and human perception based on the task taxonomy for robotic manipulators concluded by [83].

2.4 Hypotheses

Based on prior literature, we outline several hypotheses to address the research questions.

Previous research highlights that implicitly conveying contextual information through language can foster mutual understanding and facilitate smoother teamwork [25, 37]. Therefore, we hypothesise that enabling the robot to understand ISAs will positively influence team fluency and goal alignment.

H1a Perceived team fluency will be better when the robot has the capability to understand ISAs.

H1b Perceived goal alignment will be better when the robot has the capability to understand ISAs.

Existing literature suggests that ISAs can increase trustworthiness in social interaction scenarios [79]. However, there is a lack of research on their impact in physical collaborative scenarios. In this study, we hypothesise that a robot's ability to understand ISAs will positively impact trust in physical collaboration contexts.

H2 Participants will perceive the robot as more trustworthy when it has the capability to understand ISAs.

Research shows that more human-like robots are perceived as better teammates [110]. Moreover, human-like communication has been statistically proven to be highly effective in enhancing the impact of anthropomorphism compared to other anthropomorphic morphologies [70]. Thus, we hypothesise that the robot's ability to understand ISAs will enhance the user experience by increasing its perceived anthropomorphism, making interactions feel more human-like.

H3 Participants will perceive the robot as exhibiting greater anthropomorphism when it has the capability to understand ISAs.

3 User Study

To investigate the impact of a robot's ability to understand indirect speech acts on people's perception, we conducted a Wizard-of-Oz experiment with 36 participants on three different physical collaborative tasks.

The experiment employed a mixed-method experimental design [28], collecting quantitative data through a questionnaire on team fluency, goal alignment, performance trust, and anthropomorphism as dependent variables, as well as qualitative data from interview responses. The Speech Mode (ISA vs. Non-ISA) served as a between-subject factor, with half of the participants interacting with a robot capable of understanding ISAs, while the other half interacted with a robot unable to comprehend ISAs. Each participant completed three tasks in counter-balanced order with the robot using one of the assigned Speech Modes, which was followed by a semi-structured interview. We provide additional detail on our experimental design in the following sections.

3.1 Experimental Design

3.1.1 Apparatus and Setup. We used TIAGo as the robot agent in our study. TIAGo is a mobile manipulator robot with anthropomorphic features, including a head, neck, torso, and arm, making it well-suited for HRI research [68]. The robot and the participant were on the opposite side of a table, which acted as the shared workspace between the two parties. Given the current limitations of algorithms in achieving human-level understanding and generating accurate verbal responses to ISAs, and to minimise the influence of potential robotic failures on experimental outcomes, we chose to use a Wizard-of-Oz (WoZ) approach. WoZ is a classic methodology in HRI research, where human operators discretely control the robot's behaviour to simulate advanced robotic capabilities that the system itself may not be able to achieve autonomously yet or that would not be robust for real-time interaction [56]. This approach allows researchers to focus on understanding user interactions with the robot without being hindered by technological limitations or safety issues related to autonomous motion control. In this experiment, two experimenters discreetly controlled the robot to provide realistic and fluid responses, enabling a better assessment of human-robot interaction dynamics.

One of the experimenters (Motion Wizard) was teleoperating the robot's movement behind a one-side mirror, which allowed them to have a clear view of both the robot and the participant while remaining hidden from the participant. This teleoperation was possible thanks to custom-made software developed by our team that allowed the Motion Wizard to send commands to the robot remotely. This WoZ software was built using the Robotics Operating

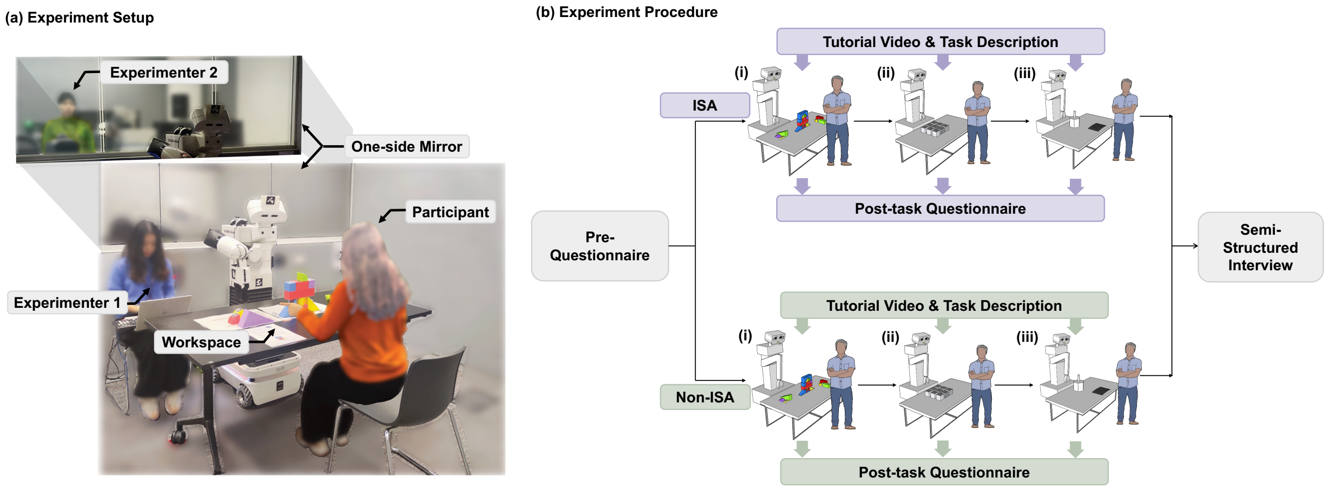


Figure 2: (a) Experiment setup: The participant, robot, and experimenter 1 were all present in the same room. The participant and robot were seated on opposite sides of a table, with the shared workspace located in the centre. Experimenter 1 (Speech Wizard) sat next to the robot, near the emergency button, and operated the robot's speech-WoZ interface. Experimenter 2 (Motion Wizard) was positioned behind a one-side mirror, allowing for a clear view of the room, and was responsible for teleoperating the robot's arm movements. (b) Experiment procedure: Each participant first completed a pre-questionnaire before being assigned to either the ISA or non-ISA group. The participant then performed three tasks with the robot in a counter-balanced order. Before each task, participants watched a tutorial video and read a task description. After completing each task, they filled out a post-task questionnaire. The experiment concluded with a semi-structured interview.

System (ROS) and the TIAGo API. We implemented the arm actions using inverse kinematics, which calculates the joint configuration based on the desired Cartesian coordinates of the end effector [22]. In addition to moving the end effector within a 3D space above the workspace, the robot's head also had 2 degrees of freedom, which allowed the Motion Wizard to observe through the robot's camera and actively engage with the participant. Safety was ensured by a collision detection function that automatically disabled the arm controller when abnormal tolerance values were detected in the joints. Virtual walls were also implemented around the robot's arm to restrict its movement, preventing it from exceeding a designated range or approaching the participant too closely.

The other experimenter (Speech Wizard) was sitting beside the robot's emergency button and operating the speech-WoZ interface through a laptop to give verbal responses, which were scripted in advance (See subsubsection 3.1.2 in detail). Participants were informed that the Speech Wizard served as a safeguard, responsible for ensuring their physical safety by using the emergency button located on the robot's base if necessary. This explanation led participants to view the Speech Wizard's presence as a precautionary measure. TIAGo utilises Acapela Group's Text-to-Speech technology, which carries out the phonetic transcription of the text, generates prosody for the speech, produces the audio signal, and plays through TIAGo's speaker. Figure 2a demonstrates the experimental setup.

3.1.2 Robot's Speech Understanding. The robot's Speech Mode was a between-subject independent variable with two conditions. In the ISA condition, the robot could understand participants' ISAs and respond with appropriate actions. In the Non-ISA condition, the robot was only able to grasp the literal meaning of requests and respond to commands that were stated in imperative sentences. The literal meaning of ISAs was interpreted by isolating them from their contextual elements, following the guidelines of Searle's putative facts [82]. We selected some representative requests from participants to demonstrate how the direct and indirect speech acts were interpreted and responded to during our experiment (shown in Table 1 and Figure 1). To respond to both indirect and direct requests, the speech-WoZ interface featured predefined sentences, such as "Sure," "Okay, working on that," and "Yes, I have the ability to do that." Utterances without command intent, such as "Thank you, TIAGo," were responded to as natural conversational exchanges like "You're welcome." The selection of phrases was guided by the aforementioned literature and further refined through insights gained from four pilot studies. The interface also provided a text box that allowed the Speech Wizard to input responses to any unexpected speech. To maintain the flow of interaction and avoid constraining the use of ISAs, participants were allowed to use gestures along with their speech, which experimenters interpreted and responded to accordingly. Notably, no participants reported noticing that the experimenter sitting in front of them was controlling the robot's speech.

3.1.3 Collaboration Tasks. A recent systematic review [83] categorised the HRC tasks for robotic manipulators as: (1) collaborative assembly, where humans and robots work together to assemble complex objects through a series of sequential sub-processes; (2) object handling & handover, involving the joint grasping and placement

Table 1: Examples of participants' requests, interpretations, and robot's responses in different Speech Modes. The request examples are from Figure 1. (P: participant; R: robot)

Request Examples	Interpretations	Robot's Responses	
		If in the ISA group	If in the Non-ISA group
P: Please also sort these cubes.	**Direct** *Literal*: Sort cubes. *Intent*: Sort cubes.	R: Yes, sure. (Act on the intent)	
P: Can you move it to here please?	**Indirect** *Literal*: Ask for the ability to move it. *Intent*: Move it.	R: Got it. (Act on the intent)	R: Yes, I can do that. (No action)
P: You could take the small yellow cylinder ...	**Indirect** *Literal*: Suggest an action option to take the cylinder. *Intent*: Take the cylinder.	R: Okay. (Act on the intent)	R: Well noted. (No action)
P: Let's move on to the third column.	**Indirect** *Literal*: Suggest moving on to the third column. *Intent*: Sort the third column.	R: Working on that. (Act on the intent)	R: It's a good suggestion. (No action)
P: And the last one.	**Indirect** *Literal*: A reference to the last thing. *Intent*: Rotate to the last face.	R: Sure. (Act on the intent)	R: ... (Silence. No action)
P: The blue block needs to ... P: Oh, actually. Wait.	**Indirect** *Literal*: Provide information for the blue block and wait. *Intent*: Move the blue block to a position. The last command is wrong, stop the current action and wait for the next one.	R: Got it. (Act on the intent, then stop halfway)	R: Thank you for the info. (No action)

of objects by humans and robots, as well as the handover of objects from the robot to the human; and (3) collaborative manufacturing, where both humans and robots perform tasks that permanently alter an object, such as polishing and drilling. For safety considerations, we modified the object handling and handover task to a turn-based pick-and-place activity, where both the robot and the human participated in sorting cubes. Based on this taxonomy, we designed and implemented three physical collaborative tasks for our experiment: (1) a foam brick assembly task [101], (2) a 3*3 cubes sorting task [35], and (3) a hexagonal prism polishing task [63]. In each task, the robot lacked prior information about the task's goal and plan, requiring the participant to relay the instructions to the robot at the beginning and verbally guide the team's actions throughout the entire activity.

The **assembly** task (Figure 2bi) required the human-robot team to build a structure using foam bricks. We distributed 18 bricks of various shapes between the human and robot, with 12 bricks required for constructing the target structure and 6 incorrect bricks that should not be used. Only the participant was provided with a photo of the structure they had to build, while the robot had no prior knowledge of the structure. The bricks were initially randomly placed in the robot or the participant's stock, and each party was only allowed to take bricks from their own pile. Participants could only manipulate the bricks on their side and needed to communicate and coordinate with the robot to have it add its bricks to the construction.

The **sorting** task (Figure 2bii) used nine 5*5*5cm cubes that needed to be rearranged according to two categorical attributes: texture and version. Each cube featured a type of surface texture (smooth, medium, rough), and an ArUco marker [39] encoding its version information (old, intermediate, new). To mimic an information asymmetry sorting task, the participants were able to touch and feel the texture, whereas the robot could scan the ArUco marker to access the cube's version. We used apparently similar ArUco markers for version information to make it impossible for participants to distinguish between them by sight alone. Only the exchange of information between teammates made it possible to achieve the task: to arrange the cubes in a gradient from rough to smooth in one dimension and from old to new in the orthogonal direction.

The **polishing** task (Figure 2biii) constituted a simple instantiation of a manufacturing task. The robot was responsible for holding and turning the hexagonal prism, and the participant polished each surface three times using sandpaper. Every time the participants were happy with the sanding, they had to communicate to the robot to turn the object to show a face that had not been polished. This scenario was designed to simulate a situation where the hexagonal prism was too heavy or hazardous for a human to lift and rotate, requiring cooperation with the robot to successfully complete the task.

3.2 Participants

We conducted *a priori* power analysis to calculate the sample size for our experiment using *G*Power* [36]. The calculation was based on a medium effect size of $f = 0.25$, an alpha-level of 0.05, and a power of 0.9. As a result, we recruited 36 ($Female : Male = 19 : 17$, $M_{age} = 24.08$, $Std_{age} = 5.75$) participants who were all fluent English speakers. We used the three questions of the interaction subscale from the Negative Attitude Toward Robots Scale (NARS Questionnaire) [65], as they were relevant to working and talking to a robot (see Appendix A). These questions were used to screen out individuals who exhibited strong negative responses towards robots and who could possibly feel distressed interacting with a robot (i.e., a rating higher than 3). Each experiment took about 60 minutes and participants were compensated with a $30 voucher. Our experiment received ethics approval from the Institutional Review Board (IRB).

3.3 Procedure

The experiment procedure is shown in Figure 2b. Upon welcoming the participants, the study started with a pre-questionnaire, which captured participant demographics and their prior interaction experience with robots, voice assistants, and in performing physical collaborative tasks. The prior experience served as covariates in data analysis. Before each task, participants were provided with a tutorial video and a written task description, which included instructions and specified the objectives of the task. Additionally, a picture of the target structure for the assembly task was presented to illustrate the final goal. Participants were required to lead the collaboration and verbally relay the team's objectives to their robot teammate, TIAGo. After each task, participants completed a post-task questionnaire that assessed their perceptions of the team's fluency and goal alignment [41], performance trustworthiness using Multi-Dimensional Measure of Trust (MDMT) [54], and the robot's anthropomorphism using the Godspeed Questionnaire (GSQ) [6]. Each participant interacted with one of the robot's Speech Modes (ISA or Non-ISA) and engaged in three tasks, which were assigned in a counter-balanced order. Finally, the study ended with a semi-structured interview. Each experiment took about 60 minutes, including the interview.

3.4 Data Collection and Analysis

We collected the quantitative data using standard questionnaires and the qualitative data through a semi-structured interview. The following dependent variables were collected after each task:

- Team fluency: To answer RQ1.1, we used the 7-point team fluency sub-scale with 3 items, from [41], which adapted the Working Alliance Inventory [42] on Human-Robot Collaboration.
- Goal alignment: For the goal alignment in RQ1.2, we utilised the 7-point goal sub-scale with 3 items, from [41].
- Performance trust: The 4-item MDMT performance trust scale results were collected to measure participants' perceived capability and reliability of the robot (RQ2). The scale has 5 points and an additional option for "Does not fit" to prevent forced and possibly meaningless ratings [54].
- Anthropomorphism: To answer RQ3, the 5-point anthropomorphism sub-scale with 5 items of the GSQ was used. As the study was focused on the robot's understanding of communication rather than the appearance of the robot, the last item, "Moving rigidly/elegantly", was changed to "communicating rigidly/elegantly", which has been shown to be reliable by [46].

To analyse the impact of the robot's Speech Modes (ISA vs. Non-ISA) and covariates (participants' prior interaction experience with robots, voice assistants, and physical collaborative tasks), we used Cumulative Link Mixed Models (CLMMs) via the "ordinal" package in R [23]. This analysis is appropriate given the ordinal nature of our dependent variables. Additionally, task type, scales' sub-item ID, and participant ID were included as random effects in our model to account for potential variability within group structures and repeated measures [18].

At the end of the experiment, we conducted a semi-structured interview lasting approximately 15 minutes to gather qualitative feedback from participants. The Motion Wizard observed participants' behaviours during the experiment. Instances of participants using indirect speech acts were further explored through follow-up questions during the interviews. The interviews were intended to supplement the quantitative results and provide insight into their subjective feelings regarding the overall experience during the collaboration. Given the between-subjects design of the study, we began the interview by explaining the experimental condition that participants had not experienced, ensuring they had a comprehensive understanding of the study. We disclosed that the experimenters controlled the robot's actions and speech only after the interview concluded.

The interview results were transcribed and analysed through reflexive thematic analysis (RTA), which was well-suited to this study because it emphasised the researchers' active role in constructing themes, thereby fostering flexibility, creativity, and critical reflection. This approach permits researchers to integrate their own insights and observations from the experimental process, making it particularly effective for exploring subtle phenomena [13]. Following the 6-phase guidance by [11], two authors of this paper, both of whom possess substantial expertise in human-robot interaction and human-computer interaction, conducted the RTA. In phase 1, researchers thoroughly reviewed all transcriptions. In phase 2, they inductively generated initial codes at the sentence level, which were either semantic, representing participants' explicit feelings, or latent, reflecting deeper meanings inferred from the data based on researchers' knowledge background. In phase 3, they constructed the initial themes and categorised the codes. Up to this point, the work had been carried out individually by each researcher. In phase 4, two researchers cooperatively discussed and reviewed the themes through multiple rounds. In phase 5, the themes were defined and named. In phase 6, researchers drafted the initial report of the qualitative analysis. Phases 4 to 6 were repeated over several rounds, during which the themes were iteratively refined and discrepancies addressed. This process aligns with the RTA principles, which emphasise continuous iterative reflexivity to ensure the analysis remains progressively recursive [97].

4 Results

Table 2 shows a summary of participants' demographics and their prior interaction experience with robots, voice assistants, and physical collaborative tasks (with either humans or robots). Next, we report our quantitative and qualitative findings.

4.1 Quantitative Findings

In this section, we present the key results from CLMM analysis. Detailed results regarding covariates, random effects, model fit, and model formula are provided in Appendix B.

4.1.1 RQ1.1: How does a robot's capability to understand indirect speech acts influence the fluency of human-robot teamwork? Participants in the ISA group reported significantly greater perceptions of team fluency compared to those in the Non-ISA group ($\beta = 0.961, SE = 0.403, z = 2.382, p = 0.017$), as seen in Table 3. Therefore, H1a is confirmed. Figure 3a illustrates the distribution of participants' responses. The team fluency questionnaire was consistent and reliable (Cronbach's $\alpha = 0.801$) [40].

4.1.2 RQ1.2: How does a robot's capability to understand indirect speech acts influence the establishment of goal alignment among the human-robot team? We observed that the Speech Mode had a significant impact on goal alignment, with participants in the ISA group expressing a stronger belief that they were working toward a mutual goal with the robot ($\beta = 2.309, SE = 0.656, z = 3.518, p < 0.001$). Therefore, H1b is confirmed. This is further illustrated in Figure 3b, which shows their scaled responses.

Moreover, participants' prior experience significantly influenced their perception of goal alignment. Namely, those with greater experience in physical collaborative tasks provided significantly higher scores on the goal alignment scale ($\beta = 0.536, SE = 0.270, z = 1.985, p = 0.047$). The goal alignment questionnaire was also consistent and reliable (Cronbach's $\alpha = 0.794$) [40].

4.1.3 RQ2: How does a robot's capability to understand indirect speech acts influence a human teammate's trust in the robot's performance? The ISA group demonstrated significantly higher trust in the robot's performance compared to the Non-ISA group ($\beta = 1.105, SE = 0.493, z = 2.240, p = 0.025$). Therefore, H2 is confirmed. The MDMT performance trust questionnaire was consistent and reliable (Cronbach's $\alpha = 0.92$) [100]. Detailed participant responses can be seen in Figure 3c.

4.1.4 RQ3: How does a robot's capability to understand indirect speech acts influence a human teammate's perception of the robot's anthropomorphism? Figure 3d shows participants' responses to perceiving the robot's anthropomorphism under different Speech Modes. As shown in Table 3, participants in the ISA condition exhibited significantly higher perceptions of the robot's anthropomorphism compared to those in the Non-ISA condition ($\beta = 2.708, SE = 0.674, z = 4.016, p < 0.001$). Therefore, H3 is confirmed. The anthropomorphism sub-scale of the GSQ questionnaire was consistent and reliable (Cronbach's $\alpha = 0.94$) [46].

4.1.5 Summary. Overall, a robot's ability to understand indirect speech acts significantly influences human perception of teamwork fluency, goal alignment, performance trust, and robot anthropomorphism. Regarding the covariates, only participants' prior experience with physical collaborative tasks has a significant positive influence on the human-robot team's goal alignment. Moreover, the Speech Mode has a higher effect on goal alignment and anthropomorphism, followed by a medium effect on performance trust and team fluency.

4.2 Qualitative Findings

Two researchers thoroughly analysed participants' 502-minute semi-structured interview recordings [12]. Additionally, observations made by the experimenters during the experiment that were relevant to the interview findings were also analysed. In the following sections, we present the themes derived from interview responses.

4.2.1 Reasons for using (In)direct requests. The most frequently mentioned reason for using indirect requests during collaboration was politeness. Participants preferred to use *"Can you ...?"*—the conventionalised ISA—to show politeness in requests. They believed this approach followed social norms and felt more natural and comfortable. $P4_{Non-ISA}$ also mentioned that using ISA could offer the teammate the option to reject the request. However, $P3_{Non-ISA}$ disagreed, believing it unnecessary to be polite to a robot, which saved their effort. Therefore, she preferred to use direct commands when interacting with robots. Additionally, participants from the Non-ISA group highlighted that even after realising the robot could not understand their indirect requests, they sometimes inadvertently used ISA because it was natural and subconscious.

Due to the subconscious nature, participants who converted indirect requests to direct ones noted using ISA caused less cognitive workload. $P4_{Non-ISA}$ and $P6_{ISA}$ analogised this conversion process to constructing prompts, which requires additional time and effort. However, it was unnatural and more challenging to formulate prompts mentally and articulate them verbally when facing a physical entity, whether it was a robot or a human teammate. Furthermore, $P4_{Non-ISA}$ emphasised that if there were multiple human teammates and a robot teammate in one group, it added unnecessary mental load to switch between direct and indirect communication.

> $P4_{Non-ISA}$: "[On using ChatGPT] It didn't feel like talking to an actual figure. It did make you feel a bit more like I'm not even asking you; I'm just telling you to do something, which just didn't seem natural to me in speech form. But if it was typed out, it would be a bit easier to do that. But I have to say I have to process it a little bit more. If it's just a text screen, then I feel like there's less of a need to express any of that [ISAs] through text form or if it's just like a virtual assistant."

Moreover, participants' expectation of the robot's capability influenced their communication strategies. Those who believed the robot had a high level of understanding were more likely to use indirect commands ($P1_{ISA}$, $P15_{ISA}$). Conversely, $P27_{ISA}$ predominantly used direct requests, despite being in the ISA group, as his extensive experience with LLMs led him to doubt the AI's ability to understand implicit requests. He believed direct commands were crucial for successful task completion, even though this approach required more effort to construct explicit commands mentally. Interestingly, participants had differing perceptions regarding the simplicity of

Table 2: Overview of participants ' demographic information and their prior interaction experience with robots, voice assistants, and physical collaborative tasks. (Rarely: less than once a month; Sometimes: at least once a month but less than once a week; Often: at least once a week but less than once a day; Very often: at least once a day.)

Gender		Age		Robots		Voice Assistant		Physical Collaborative Tasks	
Female	52.8%	18-25	72.2%	No experience	33.3%	Never	5.6%	Never	25.0%
Male	47.2%	26-35	19.4%	With domestic robot	55.6%	Rarely	55.6%	Rarely	22.2%
		36-45	8.3%	With desktop pet robot	2.8%	Sometimes	19.4%	Sometimes	27.8%
				With social robot	0.0%	Often	13.9%	Often	19.4%
				With industrial robot	8.3%	Very often	5.6%	Very often	5.6%
				With more than one type	22.2%				

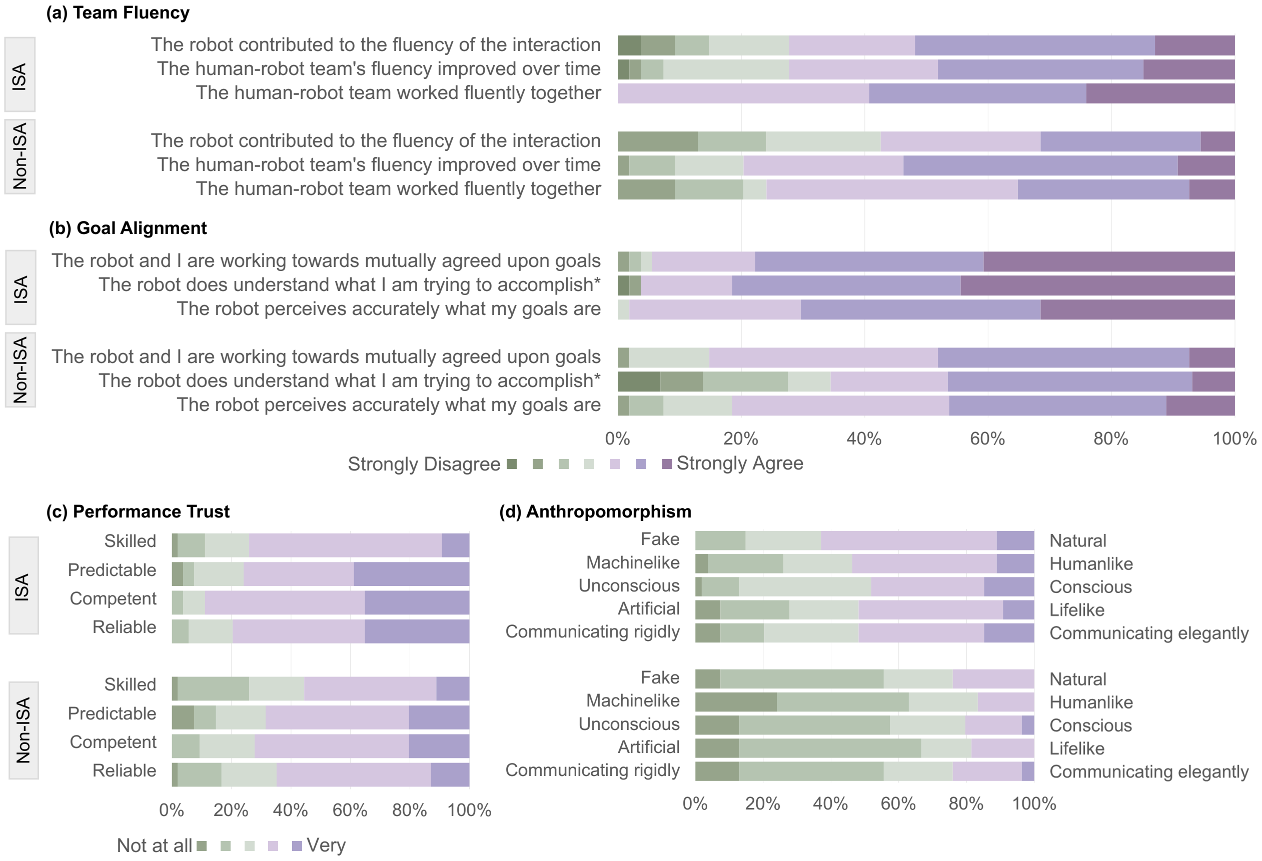


Figure 3: Participant responses on their perceptions of the team fluency (a), goal alignment (b), performance trust (c), and the robot's anthropomorphism (d) under different Speech Modes. (*: This item was originally an inverse item according to [41]. To make this figure look consistent, we reversed this item ($current\_score = 8 - original\_score$).

commands. $P2_{ISA}$ believed that direct commands were simpler for both humans and robots. In contrast, participants stated that using indirect requests felt simpler and intuitive because it was speaking aloud what was already in mind. *"Like an extension [of mind]"*, said $P13_{ISA}$.

However, there were several differing opinions on direct requests. The most frequently mentioned reason for using direct commands was clarity. Participants believed that direct commands were better suited for tasks requiring precise and nuanced descriptions, whereas indirect requests were more likely to cause ambiguity and misunderstandings. $P5_{Non-ISA}$ and $P17_{Non-ISA}$ further explained that communication strategies exhibit task dependency. For high-risk tasks, direct commands are preferred because unambiguous instructions are critical. In contrast, more complex but low-risk tasks, as well as those requiring intensive collaboration, benefit from indirect and natural communication.

Table 3: The key results from CLMM analysis. Italics are covariates.

	Fixed Effects	Estimates	Std Error	95% CI	z	p-value
Team Fluency	Speech Mode (Non-ISA)	0.961	0.403	0.17 – 1.751	2.382	0.017*
	Robot (No)	0.004	0.110	-0.211 – 0.218	0.032	0.974
	Voice Assistant (Never)	0.129	0.210	-0.283 – 0.541	0.615	0.538
	Physical Collaborative Tasks (Never)	0.193	0.174	-0.147 – 0.533	1.112	0.266
Goal Alignment	Speech Mode (Non-ISA)	2.309	0.656	1.023 – 3.596	3.518	<0.001***
	Robot (No)	0.120	0.170	-0.214 – 0.453	0.701	0.483
	Voice Assistant (Never)	-0.099	0.316	-0.719 – 0.521	-0.312	0.755
	Physical Collaborative Tasks (Never)	0.536	0.270	0.007 – 1.064	1.985	0.047*
Performance Trust	Speech Mode (Non-ISA)	1.105	0.493	0.138 – 2.072	2.240	0.025*
	Robot (No)	-0.041	0.136	-0.307 – 0.226	-0.298	0.766
	Voice Assistant (Never)	0.231	0.248	-0.255 – 0.717	0.932	0.351
	Physical Collaborative Tasks (Never)	0.400	0.211	-0.014 – 0.814	1.892	0.058
Anthropomorphism	Speech Mode (Non-ISA)	2.708	0.674	1.387 – 4.03	4.016	<0.001***
	Robot (No)	-0.168	0.184	-0.528 – 0.192	-0.915	0.360
	Voice Assistant (Never)	0.031	0.340	-0.635 – 0.697	0.092	0.927
	Physical Collaborative Tasks (Never)	-0.111	0.288	-0.676 – 0.454	-0.385	0.701

4.2.2 *Adaptation in team fluency.*

Participants in the ISA group believed the robot's ability to understand ISA contributed to a higher team fluency. $P1_{ISA}$ and $P18_{ISA}$ agreed indirect commands enabled more flexibility in communication. However, participants in the Non-ISA group reported feeling halted, but most of them further added that it wouldn't be a problem once they adapted. Observations by the experimenters revealed that participants in the Non-ISA group often did not immediately recognise that the issue was due to the misunderstanding of their intentions. Instead, they believed it was a voice recognition problem. As a result, they tended to repeat their indirect requests slowly and word by word, which repeatedly interrupted the collaborative process. Over time, once participants in the Non-ISA group understood that the robot only responded to direct requests, they began using direct commands more consistently, although they occasionally reverted to indirect requests due to the subconsciousness, as discussed in section 4.2.1.

According to $P4_{Non-ISA}$, $P8_{Non-ISA}$, and $P11_{Non-ISA}$, although direct requests caused more mental work and initially affected team fluency, they believed this issue would diminish once humans adapted to the robot's communication abilities. *"I think once you get used to it [direct requests], not so much [affects on fluency]. When I realise I have to say things in a certain way, I think it's fine. But at the start, yeah, it's a little bit off."* said $P8_{Non-ISA}$. They stated that they had no expectation for the robot to adapt to human communication style, as it was human's responsibility to ensure the robot could understand their instructions. However, not all the users were able to adapt. Despite being aware of the robot's limitations, $P16_{Non-ISA}$ still preferred to use indirect commands.

4.2.3 *Grounding and goal alignment.*

Indirect commands allowed for more flexibility and complexity, fostering a deeper sense of partnership and shared goals ($P9_{Non-ISA}$, $P15_{ISA}$). $P18_{ISA}$ emphasised that a common understanding of the task and goal improved seamless coordination and reduced the likelihood of misunderstandings. According to $P8_{Non-ISA}$ and $P15_{ISA}$, the robot's ability to understand and act on implicit knowledge, similar to human common sense, was crucial for human teammates. This ability included grasping context-specific cues, such as spatial description, incomplete information, and shortened sentences, without requiring detailed explanation. For instance, in the polishing task, $P7_{ISA}$ gave a shortened indirect request *"And the last one please"* to indicate that the robot should turn the hexagonal prism to the final surface, based on the prior context, *"please turn so a different surface is facing me"*. Similarly, in the assembly task, $P13_{ISA}$ used an indirect request with ambiguous information, *"Oh, actually, wait"*, to signal the robot to stop its current movement, with prior context information *"Next, this blue block here needs to go up here on the red block"* and follow-up information *"This red block here needs to go in the middle of this red block here"*. $P5_{Non-ISA}$ explained that he expected the robot to develop a shared understanding based on the context of his commands. As a result, he gave indirect commands, but the robot failed to interpret them correctly. These non-conventionalised ISAs were context-dependent. Our observations, as well as follow-up interview questions, revealed that participants tended to use non-conventionalised indirect commands when they were confident their intentions were aligned with the robot's understanding.

4.2.4 *Enhanced performance trust.*

Participants in the ISA group reported a high level of performance trust in the interview, which aligns with our quantitative findings. $P1_{ISA}$, $P2_{ISA}$, and $P15_{ISA}$ agreed that they perceived the robot as more capable and reliable once they recognised its ability to understand indirect commands. Some participants in the Non-ISA group believed that the capability and reliability were contingent solely on the robot's task performance rather than its communication abilities. As $P8_{Non-ISA}$ remarked, *"I think as long as there is something that I can say that will make the robot do the task, then it's still capable and reliable"*.

Additionally, $P7_{ISA}$ emphasised that perceiving the robot as more human-like could raise expectations regarding its performance and lead to potential frustration if errors occurred. Conversely, when the robot was perceived as less human-like, errors were deemed more acceptable. However, some participants also believed that applying social norms, like politeness, in their interaction humanised the robot, which sometimes led to more lenient attitudes towards errors, similar to their reactions to human teammates' mistakes.

4.2.5 *Perceptions of anthropomorphism.*

In line with the quantitative results, the majority of participants acknowledged the robot's ability to understand ISA affected their perception of the robot's anthropomorphism. They agreed that the feeling of human likeness manifested from the ability to understand, even though the voice and tone were still machine-like. $P6_{ISA}$ reported *"I think its ability to understand the implicit language made me feel like somebody was listening"*. However, some participants indicated that additional factors influenced their perceptions. $P12_{Non-ISA}$ highlighted she would *"make small talk with a human, but the robot doesn't."* Furthermore, $P25_{ISA}$ believed that the sentences used by the robot felt mechanical, which diminished his perception of the robot's human likeness.

Some participants mentioned a feeling of collaboration or control during the experiment. Participants in the Non-ISA group perceived the robot more as a tool or machine rather than a teammate ($P3_{Non-ISA}$, $P5_{Non-ISA}$, $P8_{Non-ISA}$). Participants associated this communication style with a more controlled, mechanical interaction, where the robot was seen as executing specific directives rather than participating in a collaborative process. $P8_{Non-ISA}$ and $P10_{ISA}$ agreed that direct commands required detailed instructions, which reinforced the perception of the robot as a tool needing explicit directions. This approach minimised ambiguity while simultaneously limiting the sense of shared responsibility or joint effort in the task. In this context, the robot was viewed as an extension of the user's will, carrying out predefined actions without critical decision-making. On the contrary, some participants believed the robot's ability to interpret indirect commands indicated a higher level of cognitive processing, similar to human teammates' ability to infer meaning and anticipate actions based on incomplete information. Unlike direct commands, participants perceived the robot more as a teammate rather than a tool when using indirect commands. This perception resulted from the robot's ability to understand and respond to more nuanced and context-rich communication $P1_{ISA}$, $P2_{ISA}$, $P18_{ISA}$, $P18_{ISA}$). Indirect commands were often used in a more conversational tone, suggesting a partnership where the robot was expected to understand the intent behind the instructions and act accordingly. This contributed to participants' perception of the robot's anthropomorphism ($P10_{ISA}$, $P13_{ISA}$, $P18_{ISA}$).

> $P17_{Non-ISA}$: "I think, because indirect commands and subtext is a very human-feeling thing. So when I'm just giving it a direct command, it feels more like I'm just putting an input into a machine. Whereas with the indirect commands, it feels more like I'm having a conversation with someone."

When discussing future usage, participants also expressed concerns about different aspects of the robot's anthropomorphism. Participants believed that they preferred a robot with an extremely human level of understanding but not one that mimicked human tone, voice, or appearance. $P1_{ISA}$ mentioned that it also depends on the type of task, *"If it's a vacuum cleaner, I want it to be less human cause it's just vacuuming the floor, you know. But if we're working on tasks kind of like this [our study], where it would be normal for two humans to work together, then I would definitely want the robot to be more human just so it's easier to communicate and get things done quickly."*

4.2.6 *Expectations on LLM.*

As this study used the Wizard-of-Oz method, participants assumed that the ISA robot was implemented with an LLM. Several participants compared the robot's capabilities with their prior experiences using voice assistants and commercial LLMs. $P25_{ISA}$, $P32_{Non-ISA}$, and $P36_{Non-ISA}$ believed that a novel LLM should be capable of handling indirect requests, at least the conventionalised ones, i.e. *"Can you ...?"*. However, $P25_{ISA}$ also acknowledged that he tended to be more direct when interacting with a text-based LLM. $P32_{Non-ISA}$ believed that using more indirect and polite language with ChatGPT usually yielded better outcomes. *"You have to be very patient"*, $P32_{Non-ISA}$ remarked.

In contrast, $P27_{ISA}$ explained that his experience with LLMs led him to doubt the robot's ability to comprehend indirect commands, which prompted him to provide explicit instructions to ensure task success. As a result, he primarily used direct commands during the collaboration in our study. $P4_{Non-ISA}$ and $P6_{ISA}$ concurred that ChatGPT usually performed better when using direct commands. Despite this, they both used numerous indirect commands in this study, noting that verbal commands differ from written commands as they provide less time to construct the prompts, and the formation of commands is often ad-hoc, leading to more ambiguity and incomplete sentences. Furthermore, $P6_{ISA}$, who expressed doubts about the implementation of LLMs in our robot, raised concerns about their effectiveness in real physical-embodied scenarios, arguing that LLMs would likely struggle in such contexts.

> $P6_{ISA}$: "[When interacting with the robot] I think because I'm referring to things that exist in space as opposed to a concept that exists just in our mind. So if we're talking about something like, What's the difference between a plant cell and, you know, an animal cell? It's got a text-based understanding of that. But because this [interacting with a robot] is referring to a real embodied scenario, I think it would struggle to do anything with this."

5 Discussion

In this section, we discuss the impact of indirect speech acts in human-robot collaboration based on our quantitative and qualitative findings.

5.1 ISA's subtle role in teamwork

5.1.1 *Adaptation and synchronisation.*

Participants reported that a robot capable of understanding indirect requests made fluent collaboration easier. This is consistent with our quantitative results. Moreover, participants in the Non-ISA group reported adapting to the robot's communication ability to increase team fluency over time. Previous research shows human collaborators tend to have synchronisation on their vocalisation and neural activity when selecting words to convey contextual meanings during conversations [3, 108]. Moreover, literature further suggests that people unconsciously mirror their linguistic structures with their interlocutors, regardless of being a human or computer, which facilitates efficient interactions [10]. However, the robot in the Non-ISA group failed to reciprocate linguistic convergence by adapting to their human collaborators.

Our findings revealed that the absence of robots' ability to interpret ISAs (Non-ISA group) necessitated greater adaptation efforts from participants during collaboration. This adaptation process was reported to be time-consuming and requiring increased cognitive effort. The robot's failure to perform its role as an effective communication partner forced participants to take on the full responsibility of adapting, increasing their effort and disrupting the division of labour [25, 29]. Previous studies have shown that ideal robot teammates should be able to adapt their communication to establish common ground for shared environment [21]. **Therefore, it is essential to enhance robots' ability to adapt to individuals' communication styles, for instance, using indirect requests.** Besides, to be accessible to diverse populations, robots should adapt to users' speech styles, considering factors like "age, gender, dialect, domain expertise, task knowledge, and familiarity with the robot." [55]. Our findings support this call, while highlighting the importance of indirect speech. As collaborative robots enter real-world settings, it is suboptimal to expect groups, like children or the elderly, to adapt to the robot's communication style.

5.1.2 Grounding. Previous research in human-robot interaction conducted limited exploration of non-conventionalised ISAs (i.e. context-dependent ISAs), usually focused on the effect of politeness (i.e. conventionalised ISAs) [84, 105]. In our study, we discovered significant effects of non-conventionalised ISAs on team grounding. The interview findings provided an explanation for the questionnaire results, which showed that the ISA group had a significantly higher perception of goal alignment compared to the Non-ISA group. Participants who effectively used indirect requests, particularly non-conventionalised ISAs, to communicate with the robot felt more confident in having established a shared understanding with their robot teammate. Moreover, participants mentioned that conventionalised ISAs (e.g., "Can you...?") offer the teammate an option to reject the request, consistent with Searle's [82] theory, which explains that ISAs also contribute to facilitating the exchange of intentions between teammates.

In contrast to human-human collaboration, the use of ISAs is nuanced by the users' expectations. Some participants, having no expectation of the robot's ability to understand ISAs, opted to use only direct requests, even when interacting with a robot capable of interpreting ISAs. This finding complements the results of [16], which showed that individuals continue using ISAs when interacting with robots that cannot comprehend them—a pattern also observed in our study. This shows that the user's prior expectations, or mental models, of the robot's capabilities play a strong role in people's decision to use or avoid ISAs. It could be important for a robot teammate to explicitly communicate its capabilities to interpret ISAs. At the start of a collaboration, for example, the robot might say *"Please just give me clear, precise, direct instructions"*.

With human-agent teaming on the rise, goal alignment has emerged as a critical yet unresolved challenge [8, 111]. Previous research has focused on approaches that model goal alignment and assess its effects [49, 78]. Our findings suggest that the successful use of indirect requests in communication can act as an indicator of mutual understanding within the team. Consequently, proactively **incorporating implicatures into the robot's verbal communication may be an effective strategy for signalling the robot's accurate comprehension of the human teammate's intentions**. However, ISAs can also introduce ambiguities, requiring the robot to more effectively manage dialogue failures and repair mechanisms [45]. The appropriate usage of this strategy not only enhances the explainability of the robot's mental state but also maintains the flow of teamwork without interruptions.

5.2 ISA's subtle role in trust

Qualitative findings suggest that the robot's ability to understand ISAs either positively impacted or did not affect participants' trust as long as tasks were completed successfully. This qualitative feedback supports the quantitative results, indicating that understanding ISAs significantly enhances trust, although trust remained high in the Non-ISA group due to successful task execution. Moreover, some participants felt that using indirect requests enhanced their perception of the robot's anthropomorphism. This result aligned with the findings in [33], claiming that a robot's speech anthropomorphism should not be limited to tone and voice but also to directness. Previous studies conclude that robots with higher anthropomorphism in appearance (i.e. looking more human-like) may induce higher functionality expectations [32] and trust [62]. Our study adds to these findings that higher human-like understanding in verbal communication may induce higher performance trust.

However, our study represented an ideal scenario where the robot made no mistakes. In real-world settings, execution and communication failures are common. During the interview, some participants suggested that higher performance trust could result in elevated expectations, potentially causing greater frustration when the robot makes errors. Conversely, participants with a higher perception of the robots' partnership believed they would be more forgiving of the robot's potential mistakes. Similar contradictory findings have been reported in previous research. Salem et al. [76] observed that robots displaying occasional incorrect gestures were perceived as more likeable than those that performed perfectly. A follow-up study [77] found opposing results, but also suggested that the level of anthropomorphism and the severity of the error may influence these differing reactions. In our experiment, we used a robot with anthropomorphic features, including a head, neck, arm, and torso. Some participants noted during the interviews that their perceptions might differ if the robot were less human-like, such as a vacuum robot. **Given that real-world interactions are more prone to errors, it is crucial to carefully consider the potential negative effects on trust when employing natural and implicit verbal communication in collaboration.** Although there are no widely recognized studies analysing users' speech acts when a robot fails, Kontogiorgos et al. [45] found that humans tend to emphasise vowels and speak more loudly when robots make errors. Future research could explore users' speech directness in response to robot errors, particularly in relation to the level of anthropomorphism, timing and severity of the failure [71].

5.3 Task- and context-dependency

Contrary to our assumptions, indirect requests are not suitable for all situations. The usage of ISAs is task-dependent. Participants responded that **ISAs were preferred when collaborating on**

repetitive, low- to medium-risk tasks, as well as tasks requiring high coordination. For high-risk tasks, the explicitness of direct requests is safer, as it provides clearer and more precise descriptions of the required actions. Simple and repetitive tasks typically require less verbal communication, and participants often use shortened indirect requests based on the mutual understanding of the task they have built before, such as *"Next"*. In less collaborative tasks, participants prefer fewer, clearer instructions over back-and-forth dialogue, prioritising efficiency and precision over an intuitive and low-effort interaction experience during task completion.

The use of indirect requests is also highly context dependent. Unlike written commands given to virtual AI assistants, verbal commands are given less time to formulate and are often subconsciously phrased as indirect requests during physical collaboration. Furthermore, since collaborative tasks typically involve continuous interaction and sequential sub-tasks, indirect requests often rely on prior commands and actions. Interpreting these requests requires the ability to reference previous interactions that are related. Researchers suggested ISAs are less semantically related to their immediate context than direct speech acts; however, they gain relevance when interpreted correctly in light of broader context [9]. Previous research highlights the substantial impact of incorporating task context on improving the prediction of ISA usage, emphasising the need for models that account for contextual and intentional factors [90]. In HRC, which involves real-world interactions, gestures are frequently employed, further facilitating the use of indirect commands. Additionally, implicatures often rely on real-world information, such as the location of objects. A previous study explored how locative expressions embedded in indirect commands are interpreted [47]. The physical affordances of the environment, which embed rich semantic information [17, 38], can readily prompt the use of implicatures in indirect requests. Future research could focus on developing solutions that address broader conversational contexts and link physical environments to improve the accuracy of robots' interpretation of non-conventionalised indirect requests, where large language and vision models have shown strong potential due to their long-term context-sensitive attention and multi-modal reasoning capabilities [85, 114].

5.4 Limitations and future work

With the rapid development of LLMs and enhanced reliability and affordability of robot hardware, the use of natural language as an interface for daily human-robot collaboration is becoming increasingly feasible. However, some participants noted that LLMs performance in interpreting indirect requests, based on their experiences with commercial models, varied significantly, indicating that while LLMs show some capacity for interpreting implicature, their reliability remains inconsistent. This challenge has also been highlighted and explored by other researchers [44, 73]. Moreover, the context- and task-dependent nature of using ISAs in physical human-robot collaboration presents additional challenges, particularly in integrating visual and physical information. Therefore, a future evaluation is necessary before deploying LLMs in commercial collaborative robots, along with the development of specialised datasets and fine-tuning techniques. There is also potential for expanding the scope to broader conversational contexts and linking physical environments to enhance robots' interpretation of non-conventionalised indirect requests, where recent advances in language and vision models offer promising solutions with their context-sensitive attention and multi-modal reasoning capabilities.

There are several limitations in our study. First, although participants represented a wide age range, most were recruited from a university campus, with the majority being college students, which limits the generalisability of our findings to the broader population or specific demographic groups. Second, due to the use of teleoperation to ensure safety, participants noted that the robot's arm movements were slow and unsteady, which may have influenced their perception of the robot's capabilities. Third, this experiment employed a Wizard-of-Oz setup, which created an error-free scenario, allowing us to focus on analyzing users' behaviour in using ISAs and comparing across Speech Modes. However, robot errors are unavoidable in real-world applications. Future research should investigate how robot errors influence the directness of user speech, as well as their impact on collaboration performance and experience. Fourth, this experiment did not control the amount of ISAs each participant used during their interaction session to maintain natural interactions. The effect of this variation was considered in a broader sense by accounting for participants and task type as random effects rather than precisely measuring the number of ISAs used. Additionally, this study exclusively examined the impact of robots' ability to understand ISAs. Future research should explore the effects of robots' ability to generate ISAs in collaborative settings. Another important area for investigation is the influence of robots using ISAs on human communication patterns, particularly whether a robot's use of ISAs encourages humans to reciprocate with indirect speech. This dynamic could impact users' tolerance for dialogue errors, with effective ISA use potentially fostering greater flexibility and tolerance for minor mistakes, while improper handling of ISAs could reduce trust and interaction fluency.

6 Conclusion

In this study, we investigated the impacts of indirect speech acts on human-robot collaboration. Our findings highlight that the robot's ability to interpret ISAs plays a crucial role in verbal communication, though the implications of this ability vary depending on context and task. Our results suggest that ISAs hold significant potential as a communication tool to facilitate team fluency, goal alignment, and trust in HRC when applied appropriately. Robots with the ability to understand indirect requests can also increase human perception of anthropomorphism, which enhances the sense of partnership and results in a better collaborative experience. We further explored the human motivations for using indirect requests and the underlying factors driving these impacts using qualitative analysis.

Future research should focus on assessing language models' ability to interpret implicatures in indirect requests, provide appropriate ISAs, and develop large language models capable of nuanced, context-aware interactions for robotic systems. Moreover, given the inherent ambiguity of ISAs, designing effective backchanneling mechanisms to prevent misunderstandings and convey uncertainty is equally important. We advocate for careful integration of both direct and indirect verbal communication into the design and evaluation of collaborative robots, ensuring that ISAs are neither overlooked nor overused in inappropriate contexts.

Acknowledgments

We gratefully acknowledge the support provided by the Melbourne Research Scholarship (University of Melbourne) and the Australian Research Council Discovery Early Career Research Award (Grant No. DE210100858).

References

[1] Hussein A Abbass, Jason Scholz, and Darryn J Reid. 2018. *Foundations of trusted autonomy.* Springer Nature. doi:10.1007/978-3-319-64816-3

[2] Sameera A Abdul-Kader and John C Woods. 2015. Survey on chatbot design techniques in speech conversation systems. *International Journal of Advanced Computer Science and Applications* 6, 7 (2015). doi:10.14569/IJACSA.2015.060712

[3] Drew H Abney, Alexandra Paxton, Rick Dale, and Christopher T Kello. 2021. Cooperation in sound and motion: Complexity matching in collaborative interaction. *Journal of Experimental Psychology: General* 150, 9 (2021), 1760. doi:10.1037/xge0001018

[4] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N Bennett, Kori Inkpen, et al. 2019. Guidelines for human-AI interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems.* Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3290605.3300233

[5] Peter Anderson, Qi Wu, Damien Teney, Jake Bruce, Mark Johnson, Niko Sünderhauf, Ian Reid, Stephen Gould, and Anton Van Den Hengel. 2018. Vision-and-language navigation: Interpreting visually-grounded navigation instructions in real environments. In *Proceedings of the IEEE conference on computer vision and pattern recognition.* IEEE, 3674–3683. https://openaccess.thecvf.com/content_cvpr_2018/html/Anderson_Vision-and-Language_Navigation_Interpreting_CVPR_2018_paper.html

[6] Christoph Bartneck. 2023. Godspeed questionnaire series: Translations and usage. In *International Handbook of Behavioral Health Assessment.* Springer, 1–35. doi:10.1007/978-3-030-89738-3_24-1

[7] Maxwell Bennett, Tom Williams, Daria Thames, and Matthias Scheutz. 2017. Differences in interaction patterns and perception for teleoperated and autonomous humanoid robots. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).* IEEE, 6589–6594. doi:10.1109/IROS.2017.8206571

[8] Shreyas Bhat, Joseph B Lyons, Cong Shi, and X Jessie Yang. 2024. Value alignment and trust in human-robot interaction: Insights from simulation and user study. In *Discovering the Frontiers of Human-Robot Interaction: Insights and Innovations in Collaboration, Communication, and Control.* Springer, 39–63. doi:10.1007/978-3-031-66656-8_3

[9] Isabella P Boux, Konstantina Margiotoudi, Felix R Dreyer, Rosario Tomasello, and Friedemann Pulvermüller. 2023. Cognitive features of indirect speech acts. *Language, Cognition and Neuroscience* 38, 1 (2023), 40–64. doi:10.1080/23273798.2022.2077396

[10] Holly P Branigan, Martin J Pickering, Jamie Pearson, and Janet F McLean. 2010. Linguistic alignment between people and computers. *Journal of pragmatics* 42, 9 (2010), 2355–2368. doi:10.1016/j.pragma.2009.12.012

[11] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101. doi:10.1191/1478088706qp063oa

[12] Virginia Braun and Victoria Clarke. 2012. *Thematic analysis.* American Psychological Association. doi:10.1037/13620-004

[13] Virginia Braun, Victoria Clarke, Nikki Hayfield, Louise Davey, and Elizabeth Jenkinson. 2023. Doing reflexive thematic analysis. In *Supporting research in counselling and psychotherapy: Qualitative, quantitative, and mixed methods research.* Springer, 19–38. doi:10.1007/978-3-031-13942-0_2

[14] Cynthia Breazeal. 2004. *Designing sociable robots.* MIT press. doi:10.7551/mitpress/2376.001.0001

[15] Gordon Briggs and Matthias Scheutz. 2013. A hybrid architectural approach to understanding and appropriately generating indirect speech acts. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 27. AAAI Press, 1213–1219. doi:10.1609/aaai.v27i1.8471

[16] Gordon Briggs, Tom Williams, and Matthias Scheutz. 2017. Enabling robots to understand indirect speech acts in task-based interactions. *Journal of Human-Robot Interaction* 6 (2017), 64–94. doi:10.5898/JHRI.6.1.Briggs

[17] Anthony Brohan, Yevgen Chebotar, Chelsea Finn, Karol Hausman, Alexander Herzog, Daniel Ho, Julian Ibarz, Alex Irpan, Eric Jang, Ryan Julian, et al. 2023. Do as i can, not as i say: Grounding language in robotic affordances. In *Conference on robot learning.* PMLR, 287–318. https://proceedings.mlr.press/v205/ichter23a.html

[18] Violet A Brown. 2021. An introduction to linear mixed-effects modeling in R. *Advances in Methods and Practices in Psychological Science* 4, 1 (2021), 2515245920960351. doi:10.1177/251524592096035

[19] Felix Carros, Johanna Meurer, Diana Löffler, David Unbehaun, Sarah Matthies, Inga Koch, Rainer Wieching, Dave Randall, Marc Hassenzahl, and Volker Wulf. 2020. Exploring human-robot interaction with the elderly: results from a ten-week case study in a care home. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3313831.3376402

[20] Elizabeth Cha, Anca D Dragan, and Siddhartha S Srinivasa. 2015. Perceived robot capability. In *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN).* IEEE, 541–548. doi:10.1109/ROMAN.2015.7333656

[21] Joyce Y Chai, Rui Fang, Changsong Liu, and Lanbo She. 2016. Collaborative language grounding toward situated human-robot dialogue. *ai Magazine* 37, 4 (2016), 32–45. doi:10.1609/aimag.v37i4.2684

[22] Sachin Chitta, Eitan Marder-Eppstein, Wim Meeussen, Vijay Pradeep, Adolfo Rodríguez Tsouroukdissian, Jonathan Bohren, David Coleman, Bence Magyar, Gennaro Raiola, Mathias Lüdtke, and Enrique Fernandez Perdomo. 2017. ros_control: A generic and simple control framework for ROS. *The Journal of Open Source Software* 2, 20 (Dec. 2017), 456. doi:10.21105/joss.00456

[23] Rune Haubo Bojesen Christensen. 2019. ordinal—regression models for ordinal data. *R package version* 10, 2019 (2019), 54.

[24] Herbert H Clark. 1979. Responding to indirect speech acts. *Cognitive psychology* 11, 4 (1979), 430–477. doi:10.1016/0010-0285(79)90020-3

[25] Herbert H Clark. 1996. *Using language.* Cambridge university press. doi:10.2277/0521561582

[26] Herbert H Clark and Deanna Wilkes-Gibbs. 1986. Referring as a collaborative process. *Cognition* 22, 1 (1986), 1–39. doi:10.1016/0010-0277(86)90010-7

[27] Philip R Cohen and Sharon L Oviatt. 1995. The role of voice input for human-machine communication. *proceedings of the National Academy of Sciences* 92, 22 (1995), 9921–9927. doi:10.1073/pnas.92.22.9921

[28] John W Creswell. 1999. Mixed-method research: Introduction and application. In *Handbook of educational policy.* Elsevier, 455–472. doi:10.1016/B978-012174698-8/50045-X

[29] Robert Dale and Ehud Reiter. 1995. Computational interpretations of the Gricean maxims in the generation of referring expressions. *Cognitive science* 19, 2 (1995), 233–263. doi:10.1016/0364-0213(95)90018-7

[30] Martina De Cet, Martina Cvajner, Ilaria Torre, and Mohammad Obaid. 2024. Do Your Expectations Match? A Mixed-Methods Study on the Association Between a Robot's Voice and Appearance. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces.* Association for Computing Machinery, New York, NY, USA, 1–11. doi:10.1145/3640794.3665551

[31] Wen Duan, Naomi Yamashita, Yoshinari Shirai, and Susan R Fussell. 2021. Bridging fluency disparity between native and nonnative speakers in multilingual multiparty collaboration using a clarification agent. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–31. doi:10.1145/3479579

[32] Brian R Duffy. 2003. Anthropomorphism and the social robot. *Robotics and autonomous systems* 42, 3-4 (2003), 177–190. doi:10.1016/S0921-8890(02)00374-3

[33] Cloe Z Emnett, Terran Mott, and Tom Williams. 2024. Using Robot Social Agency Theory to Understand Robots' Linguistic Anthropomorphism. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction.* Association for Computing Machinery, New York, NY, USA, 447–452. doi:10.1145/3610978.3640747

[34] Friederike Eyssel, Dieta Kuchenbrandt, Simon Bobinger, Laura De Ruiter, and Frank Hegel. 2012. 'If you sound like me, you must be more human' on the interplay of robot and user features on human-robot acceptance and anthropomorphism. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction.* Association for Computing Machinery, New York, NY, USA, 125–126. doi:10.1145/2157689.2157717

[35] Marco Faroni, Manuel Beschi, Stefano Ghidini, Nicola Pedrocchi, Alessandro Umbrico, Andrea Orlandini, and Amedeo Cesta. 2020. A layered control approach to human-aware task and motion planning for human-robot collaboration. In *2020 29th IEEE international conference on robot and human interactive communication (RO-MAN).* IEEE, 1204–1210. doi:10.1109/RO-MAN47096.2020.9223483

[36] Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. 2007. G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods* 39, 2 (2007), 175–191. doi:10.3758/BF03193146

[37] Michael C Frank and Noah D Goodman. 2012. Predicting pragmatic reasoning in language games. *Science* 336, 6084 (2012), 998–998. doi:10.1126/science.1218633

[38] Jensen Gao, Bidipta Sarkar, Fei Xia, Ted Xiao, Jiajun Wu, Brian Ichter, Anirudha Majumdar, and Dorsa Sadigh. 2024. Physically grounded vision-language models for robotic manipulation. In *2024 IEEE International Conference on Robotics and Automation (ICRA).* IEEE, 12462–12469. doi:10.1109/ICRA57147.2024.10610090

[39] Sergio Garrido-Jurado, Rafael Muñoz-Salinas, Francisco José Madrid-Cuevas, and Manuel Jesús Marín-Jiménez. 2014. Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition* 47, 6 (2014), 2280–2292. doi:10.1016/j.patcog.2014.01.005

[40] Guy Hoffman. 2019. Evaluating fluency in human-robot collaboration. *IEEE Transactions on Human-Machine Systems* 49, 3 (2019), 209–218. doi:10.1109/THMS.2019.2904558

[41] Guy Hoffman and Cynthia Breazeal. 2010. Effects of anticipatory perceptual simulation on practiced human-robot tasks. *Autonomous Robots* 28 (2010), 403–423. doi:10.1007/s10514-009-9166-3

[42] Adam O Horvath and Leslie S Greenberg. 1989. Development and validation of the Working Alliance Inventory. *Journal of counseling psychology* 36, 2 (1989), 223. doi:10.1037/0022-0167.36.2.223

[43] Razan Jaber, Sabrina Zhong, Sanna Kuoppamäki, Aida Hosseini, Iona Gessinger, Duncan P Brumby, Benjamin R Cowan, and Donald Mcmillan. 2024. Cooking With Agents: Designing Context-aware Voice Interaction. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3613904.3642183

[44] Shiyu Jin, Jinxuan Xu, Yutian Lei, and Liangjun Zhang. 2024. Reasoning grasping via multimodal large language model. *arXiv preprint arXiv:2402.06798* (2024). doi:10.48550/arXiv.2402.06798

[45] Dimosthenis Kontogiorgos, Minh Tran, Joakim Gustafson, and Mohammad Soleymani. 2021. A systematic cross-corpus analysis of human reactions to robot conversational failures. In *Proceedings of the 2021 International Conference on Multimodal Interaction*. Association for Computing Machinery, New York, NY, USA, 112–120. doi:10.1145/3462244.3479887

[46] Guy Laban and Theo Araujo. 2019. Working together with conversational agents: the relationship of perceived cooperation with service performance evaluations. In *International Workshop on Chatbot Research and Design*. Springer, 215–228. doi:10.1007/978-3-030-39540-7_15

[47] Matthew Lamm and Mihail Eric. 2017. The Pragmatics of Indirect Commands in Collaborative Discourse. In *Proceedings of the 12th International Conference on Computational Semantics (IWCS)—Short papers*. arXiv. doi:10.48550/ARXIV.1705.03454

[48] Min Kyung Lee, Sara Kiesler, and Jodi Forlizzi. 2010. Receptionist or information kiosk: how do people talk with a robot?. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*. Association for Computing Machinery, 31–40. doi:10.1145/1718918.1718927

[49] Mengyao Li and John D Lee. 2022. Modeling goal alignment in human-AI teaming: a dynamic game theory approach. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 66. SAGE Publications, 1538–1542. doi:10.1177/10711813226610

[50] Jacky Liang, Wenlong Huang, Fei Xia, Peng Xu, Karol Hausman, Brian Ichter, Pete Florence, and Andy Zeng. 2023. Code as policies: Language model programs for embodied control. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 9493–9500. doi:10.1109/ICRA48891.2023.10160591

[51] Rui Liu, Jeremy Webb, and Xiaoli Zhang. 2016. Natural-language-instructed industrial task execution. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Vol. 50084. American Society of Mechanical Engineers, V01BT02A043. doi:10.1115/DETC2016-60063

[52] Rui Liu and Xiaoli Zhang. 2019. A review of methodologies for natural-language-facilitated human–robot cooperation. *International Journal of Advanced Robotic Systems* 16, 3 (2019), 1729881419851402. doi:10.1177/1729881419851402

[53] Jacob P Macdonald, Rohit Mallick, Allan B Wollaber, Jaime D Peña, Nathan McNeese, and Ho Chit Siu. 2024. Language, Camera, Autonomy! Prompt-engineered Robot Control for Rapidly Evolving Deployment. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*. Association for Computing Machinery, New York, NY, USA, 717–721. doi:10.1145/3610978.3640671

[54] Bertram F Malle and Daniel Ullman. 2021. A multidimensional conception and measure of human-robot trust. In *Trust in human-robot interaction*. Elsevier, 3–25. doi:10.1016/B978-0-12-819472-0.00001-0

[55] Matthew Marge, Carol Espy-Wilson, Nigel G Ward, Abeer Alwan, Yoav Artzi, Mohit Bansal, Gil Blankenship, Joyce Chai, Hal Daumé III, Debadeepta Dey, et al. 2022. Spoken language interaction with robots: Recommendations for future research. *Computer Speech & Language* 71 (2022), 101255. doi:10.1016/j.csl.2021.101255

[56] Nikolas Martelaro. 2016. Wizard-of-oz interfaces as a step towards autonomous hri. In *2016 AAAI spring symposium series*. AAAI Press.

[57] Cynthia Matuszek, Evan Herbst, Luke Zettlemoyer, and Dieter Fox. 2013. Learning to parse natural language commands to a robot control system. In *Experimental robotics: the 13th international symposium on experimental robotics*. Springer, 403–415. doi:10.1007/978-3-319-00065-7_28

[58] Nikolaos Mavridis and Deb Roy. 2005. Grounded situation models for robots: Bridging language, perception, and action. In *AAAI-05 workshop on modular construction of human-like intelligence*. AAAI Press.

[59] Dipendra K Misra, Jaeyong Sung, Kevin Lee, and Ashutosh Saxena. 2016. Tell me dave: Context-sensitive grounding of natural language to manipulation instructions. *The International Journal of Robotics Research* 35, 1-3 (2016), 281–300. doi:10.1177/0278364915602060

[60] Isabela Motta and Manuela Quaresma. 2021. Users' error recovery strategies in the interaction with voice assistants (VAs). In *Congress of the International Ergonomics Association*. Springer, 658–666. doi:10.1007/978-3-030-74614-8_82

[61] Chelsea Myers, Anushay Furqan, Jessica Nebolsky, Karina Caro, and Jichen Zhu. 2018. Patterns for how users overcome obstacles in voice user interfaces. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 1–7. doi:10.1145/3173574.3173580

[62] Manisha Natarajan and Matthew Gombolay. 2020. Effects of anthropomorphism and accountability on trust in human robot interaction. In *Proceedings of the 2020 ACM/IEEE international conference on human-robot interaction*. Association for Computing Machinery, New York, NY, USA, 33–42. doi:10.1145/3319502.3374839

[63] Stefanos Nikolaidis, Ramya Ramakrishnan, Keren Gu, and Julie Shah. 2015. Efficient model learning from joint-action demonstrations for human-robot collaborative tasks. In *Proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction*. Association for Computing Machinery, New York, NY, USA, 189–196. doi:10.1145/2696454.2696455

[64] Lin Ning, Luyang Liu, Jiaxing Wu, Neo Wu, Devora Berlowitz, Sushant Prakash, Bradley Green, Shawn O'Banion, and Jun Xie. 2024. User-LLM: Efficient LLM Contextualization with User Embeddings. *arXiv preprint arXiv:2402.13598* (2024). doi:10.48550/arXiv.2402.13598

[65] Tatsuya Nomura, Tomohiro Suzuki, Takayuki Kanda, and Kensuke Kato. 2006. Measurement of negative attitudes toward robots. *Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems* 7, 3 (2006), 437–454. doi:10.1075/is.7.3.14nom

[66] Alexander Obaigbena, Oluwaseun Augustine Lottu, Ejike David Ugwuanyi, Boma Sonimitiem Jacks, Enoch Oluwademilade Sodiya, and Obinna Donald Daraojimba. 2024. AI and human-robot interaction: A review of recent advances and challenges. *GSC Advanced Research and Reviews* 18, 2 (2024), 321–330. doi:10.30574/gscarr.2024.18.2.0070

[67] Naoki Ohshima, Keita Kimijima, Junji Yamato, and Naoki Mukawa. 2015. A conversational robot with vocal and bodily fillers for recovering from awkward silence at turn-takings. In *2015 24th IEEE international symposium on robot and human interactive communication (RO-MAN)*. IEEE, 325–330. doi:10.1109/ROMAN.2015.7333677

[68] Jordi Pages, Luca Marchionni, and Francesco Ferro. 2016. Tiago: the modular robot that adapts to different research needs. In *International workshop on robot modularity, IROS*, Vol. 290.

[69] Jeba Rezwana and Mary Lou Maher. 2022. Understanding user perceptions, collaborative experience and user engagement in different human-AI interaction designs for co-creative systems. In *Proceedings of the 14th Conference on Creativity and Cognition*. Association for Computing Machinery, New York, NY, USA, 38–48. doi:10.1145/3527927.3532789

[70] Eileen Roesler, Dietrich Manzey, and Linda Onnasch. 2021. A meta-analysis on the effectiveness of anthropomorphism in human-robot interaction. *Science Robotics* 6, 58 (2021), eabj5425. doi:10.1126/scirobotics.abj5425

[71] Alessandra Rossi, Kerstin Dautenhahn, Kheng Lee Koay, and Michael L Walters. 2017. How the timing and magnitude of robot errors influence peoples' trust of robots in an emergency scenario. In *Social Robotics: 9th International Conference, ICSR 2017, Tsukuba, Japan, November 22-24, 2017, Proceedings 9*. Springer, 42–52. doi:10.1007/978-3-319-70022-9_5

[72] Andrea Ruggiero, Dominik Mahr, Gaby Odekerken-Schröder, Tiziana Russo Spena, and Cristina Mele. 2022. Companion robots for well-being: a review and relational framework. *Research handbook on services management* (2022), 309–330. doi:10.4337/9781800375659.00033

[73] Laura Ruis, Akbir Khan, Stella Biderman, Sara Hooker, Tim Rocktäschel, and Edward Grefenstette. 2023. The goldilocks of pragmatic understanding: Fine-tuning strategy matters for implicature resolution by llms. In *Proceedings of the 37th International Conference on Neural Information Processing Systems* (New Orleans, LA, USA) *(NIPS '23)*. Curran Associates Inc., Article 913, 79 pages. https://dl.acm.org/doi/10.5555/3666122.3667035

[74] Eduardo Salas, Carolyn Prince, David P Baker, and Lisa Shrestha. 1995. Situation awareness in team performance: Implications for measurement and training. *Human factors* 37, 1 (1995), 123–136. doi:10.1518/001872095779049525

[75] Roya Salehzadeh, Jiaqi Gong, and Nader Jalili. 2022. Purposeful Communication in Human–Robot Collaboration: A Review of Modern Approaches in Manufacturing. *IEEE Access* 10 (2022), 129344–129361. doi:10.1109/ACCESS.2022.3227049

[76] Maha Salem, Friederike Eyssel, Katharina Rohlfing, Stefan Kopp, and Frank Joublin. 2013. To err is human (-like): Effects of robot gesture on perceived anthropomorphism and likability. *International Journal of Social Robotics* 5 (2013), 313–323. doi:10.1007/s12369-013-0196-9

[77] Maha Salem, Gabriella Lakatos, Farshid Amirabdollahian, and Kerstin Dautenhahn. 2015. Would you trust a (faulty) robot? Effects of error, task type and personality on human-robot cooperation and trust. In *Proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction*. Association for Computing Machinery, New York, NY, USA, 141–148. doi:10.1145/2696454.2696497

[78] Lindsay Sanneman and Julie A Shah. 2023. Validating metrics for reward alignment in human-autonomy teaming. *Computers in Human Behavior* 146 (2023), 107809. doi:10.1016/j.chb.2023.107809

[79] Shane Saunderson and Goldie Nejat. 2021. Robots asking for favors: The effects of directness and familiarity on persuasive hri. *IEEE Robotics and Automation Letters* 6, 2 (2021), 1793–1800. doi:10.1109/LRA.2021.3060369

[80] Beau G Schelble, Christopher Flathmann, Nathan J McNeese, Thomas O'Neill, Richard Pak, and Moses Namara. 2023. Investigating the effects of perceived teammate artificiality on human performance and cognition. *International Journal of Human–Computer Interaction* 39, 13 (2023), 2686–2701. doi:10.1080/10447318.2022.2085191

[81] Katie Seaborn, Norihisa P Miyake, Peter Pennefather, and Mihoko Otake-Matsuura. 2021. Voice in human–agent interaction: A survey. *ACM Computing Surveys (CSUR)* 54, 4 (2021), 1–43. doi:10.1145/3386867

[82] John R Searle. 1975. Indirect speech acts. In *Speech acts*. Brill, 59–82. doi:10.1163/9789004368811_004

[83] Francesco Semeraro, Alexander Griffiths, and Angelo Cangelosi. 2023. Human–robot collaboration and machine learning: A systematic review of recent research. *Robotics and Computer-Integrated Manufacturing* 79 (2023), 102432. doi:10.1016/j.rcim.2022.102432

[84] Sukyung Seok, Eunji Hwang, Jongsuk Choi, and Yoonseob Lim. 2022. Cultural differences in indirect speech act use and politeness in human-robot interaction. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 1–8. doi:10.1109/HRI53351.2022.9889576

[85] Pierre Sermanet, Tianli Ding, Jeffrey Zhao, Fei Xia, Debidatta Dwibedi, Keerthana Gopalakrishnan, Christine Chan, Gabriel Dulac-Arnold, Sharath Maddineni, Nikhil J Joshi, et al. 2024. Robovqa: Multimodal long-horizon reasoning for robotics. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 645–652. doi:10.1109/ICRA57147.2024.10610216

[86] Julie Shah, James Wiken, Brian Williams, and Cynthia Breazeal. 2011. Improved human-robot team performance using chaski, a human-inspired plan execution system. In *Proceedings of the 6th international conference on Human-robot interaction*. Association for Computing Machinery, New York, NY, USA, 29–36. doi:10.1145/1957656.1957668

[87] Masahiro Shiomi, Takamasa Iio, Koji Kamei, Chandraprakash Sharma, and Norihiro Hagita. 2015. Effectiveness of social behaviors for autonomous wheelchair robot to support elderly people in Japan. *PloS one* 10, 5 (2015), e0128031. doi:10.1371/journal.pone.0128031

[88] Valerie K Sims, Matthew G Chin, Heather C Lum, Linda Upham-Ellis, Tatiana Ballion, and Nicholas C Lagattuta. 2009. Robots' auditory cues are subject to anthropomorphism. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 53. SAGE Publications Sage CA: Los Angeles, CA, 1418–1421. doi:10.1177/154193120905301853

[89] Ishika Singh, Valts Blukis, Arsalan Mousavian, Ankit Goyal, Danfei Xu, Jonathan Tremblay, Dieter Fox, Jesse Thomason, and Animesh Garg. 2023. Progprompt: Generating situated robot task plans using large language models. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 11523–11530. doi:10.1109/ICRA48891.2023.10161317

[90] Cailyn Smith, Charlotte Gorgemans, Ruchen Wen, Saad Elbeleidy, Sayanti Roy, and Tom Williams. 2022. Leveraging intentional factors and task context to predict linguistic norm adherence. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 44.

[91] Vasant Srinivasan and Leila Takayama. 2016. Help me please: Robot politeness strategies for soliciting help from humans. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. Association for Computing Machinery, 4945–4955. doi:10.1145/2858036.2858217

[92] Aaron St. Clair and Maja Mataric. 2015. How robot verbal feedback can improve team performance in human-robot task collaborations. In *Proceedings of the tenth annual acm/ieee international conference on human-robot interaction*. Association for Computing Machinery, New York, NY, USA, 213–220. doi:10.1145/2696454.2696491

[93] Megan Strait, Cody Canning, and Matthias Scheutz. 2014. Let me tell you! investigating the effects of robot communication strategies in advice-giving situations based on robot appearance, interaction modality and distance. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*. Association for Computing Machinery, 479–486. doi:10.1145/2559636.2559670

[94] Daniel Tanneberg, Felix Ocker, Stephan Hasler, Joerg Deigmoeller, Anna Belardinelli, Chao Wang, Heiko Wersing, Bernhard Sendhoff, and Michael Gienger. 2024. To Help or Not to Help: LLM-based Attentive Support for Human-Robot Group Interactions. *arXiv preprint arXiv:2403.12533* (2024). doi:10.48550/arXiv.2403.12533

[95] Stefanie Tellex, Ross Knepper, Adrian Li, Daniela Rus, and Nicholas Roy. 2014. Asking for help using inverse semantics. *Robotics: Science and Systems X* (2014). http://hdl.handle.net/1721.1/116010

[96] Stefanie Tellex and Deb Roy. 2006. Spatial routines for a simulated speech-controlled vehicle. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*. Association for Computing Machinery, New York, NY, USA, 156–163. doi:10.1145/1121241.1121269

[97] Gareth Terry, Nikki Hayfield, Victoria Clarke, Virginia Braun, et al. 2017. Thematic analysis. *The SAGE handbook of qualitative research in psychology* 2, 17-37 (2017), 25.

[98] Cristen Torrey, Susan R Fussell, and Sara Kiesler. 2013. How a robot should give advice. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 275–282. doi:10.1109/HRI.2013.6483599

[99] Stergiani Tsoli, Stephen Sutton, and Aikaterini Kassavou. 2018. Interactive voice response interventions targeting behaviour change: a systematic literature review with meta-analysis and meta-regression. *BMJ open* 8, 2 (2018), e018974. doi:10.1136/bmjopen-2017-018974

[100] Daniel Ullman and Bertram F Malle. 2019. MDMT: multi-dimensional measure of trust.

[101] David Vogt, Simon Stepputtis, Richard Weinhold, Bernhard Jung, and Heni Ben Amor. 2016. Learning human-robot interactions from human-human demonstrations (with applications in lego rocket assembly). In *2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids)*. IEEE, 142–143. doi:10.1109/HUMANOIDS.2016.7803267

[102] Ruchen Wen, Mohammed Aun Siddiqui, and Tom Williams. 2020. Dempster-shafer theoretic learning of indirect speech act comprehension norms. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. AAAI Press, 10410–10417. doi:10.1609/aaai.v34i06.6610

[103] Tom Williams, Gordon Briggs, Bradley Oosterveld, and Matthias Scheutz. 2015. Going beyond literal command-based instructions: Extending robotic natural language interaction capabilities. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 29. AAAI Press. doi:10.1609/aaai.v29i1.9377

[104] Tom Williams, Daniel Grollman, Mingyuan Han, Ryan Blake Jackson, Jane Lockshin, Ruchen Wen, Zachary Nahman, and Qin Zhu. 2020. "Excuse me, robot": Impact of polite robot wakewords on human-robot politeness. In *Social Robotics: 12th International Conference, ICSR 2020, Golden, CO, USA, November 14–18, 2020, Proceedings 12*. Springer, 404–415. doi:10.1007/978-3-030-62056-1_34

[105] Tom Williams, Daria Thames, Julia Novakoff, and Matthias Scheutz. 2018. " Thank You for Sharing that Interesting Fact!" Effects of Capability and Context on Indirect Speech Act Use in Task-Based Human-Robot Dialogue. In *Proceedings of the 2018 acm/ieee international conference on human-robot interaction*. Association for Computing Machinery, New York, NY, USA, 298–306. doi:10.1145/3171221.3171246

[106] Kun Xu. 2019. First encounter with robot Alpha: How individual differences interact with vocal and kinetic cues in users' social responses. *New Media & Society* 21, 11-12 (2019), 2522–2547. doi:10.1177/1461444819851479

[107] Zihao Yi, Jiarui Ouyang, Yuwen Liu, Tianhao Liao, Zhe Xu, and Ying Shen. 2024. A Survey on Recent Advances in LLM-Based Multi-turn Dialogue Systems. *arXiv preprint arXiv:2402.18013* (2024). doi:10.48550/arXiv.2402.18013

[108] Zaid Zada, Ariel Goldstein, Sebastian Michelmann, Erez Simony, Amy Price, Liat Hasenfratz, Emily Barham, Asieh Zadbood, Werner Doyle, Daniel Friedman, et al. 2024. A shared model-based linguistic space for transmitting our thoughts from brain to brain in natural conversations. *Neuron* (2024). doi:10.1016/j.neuron.2024.06.025

[109] Rui Zhang, Wen Duan, Christopher Flathmann, Nathan McNeese, Guo Freeman, and Alyssa Williams. 2023. Investigating AI teammate communication strategies and their impact in human-AI teams for effective teamwork. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (2023), 1–31. doi:10.1145/3610072

[110] Rui Zhang, Nathan J McNeese, Guo Freeman, and Geoff Musick. 2021. " An ideal human" expectations of AI teammates in human-AI teaming. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (2021), 1–25. doi:10.1145/3432945

[111] Yan Zhang. 2025. Implicit Communication of Contextual Information in Human-Robot Collaboration. *arXiv preprint arXiv:2502.05775* (2025). doi:10.48550/arXiv.2502.05775

[112] Yan Zhang, Ziang Li, Haole Guo, Luyao Wang, Qihe Chen, Wenjie Jiang, Mingming Fan, Guyue Zhou, and Jiangtao Gong. 2023. " I am the follower, also the boss": Exploring Different Levels of Autonomy and Machine Forms of Guiding Robots for the Visually Impaired. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–22. doi:10.1145/3544548.3580884

[113] Zirui Zhao, Wee Sun Lee, and David Hsu. 2023. Large language models as commonsense knowledge for large-scale task planning. In *Proceedings of the 37th International Conference on Neural Information Processing Systems* (New Orleans, LA, USA) *(NIPS '23)*. Curran Associates Inc., Red Hook, NY, USA, Article 1387, 21 pages.

[114] Wanjun Zhong, Lianghong Guo, Qiqi Gao, He Ye, and Yanlin Wang. 2024. Memorybank: Enhancing large language models with long-term memory. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. AAAI Press, 19724–19731. doi:10.1609/aaai.v38i17.29946

Appendix

A Selected Questions from the Negative Attitude Towards Robots Scale

"Please read through each of the following sentences and then indicate how frequently these statements apply to you, from (1) strongly disagree to (5) strongly agree."

- I would feel very nervous just standing in front of a robot.
- I would feel very nervous talking with robots.
- I would feel uneasy if I was given a job where I had to use robots.

B Model Results

Table A1 provides the detailed results of the quantitative analysis.

The formula we used for the CLMM model in R is:

$$dependent\_variable \sim speech\_mode + pre\_robot + pre\_va + pre\_phy + (1|task\_type) + (1|scale\_item) + (1 + task\_type|p\_id)$$

where dependent_variable represents the rating from one of the scales (team fluency, goal alignment, performance trust, and anthropomorphism); speech_mode represents the independent variable; pre_robot, pre_va, and pre_phy are covariates, representing participants' previous experience with robots, voice assistants, and physical collaborative tasks; task_type, scale_item, and p_id are random effects, representing tasks types, scales' sub-item IDs, and participant IDs.

Table A1: This table shows the results from CLMM analysis. In the column of fixed effects, italics are covariates.

Team Fluency					
Fixed Effects	**Estimates**	**Std Error**	**95% CI**	**z**	**p-value**
Speech Mode (Non-ISA)	0.961	0.403	0.17 − 1.751	2.382	0.017^{*}
Robot (No)	0.004	0.110	-0.211 − 0.218	0.032	0.974
Voice Assistant (Never)	0.129	0.210	-0.283 − 0.541	0.615	0.538
Physical Collaborative Tasks (Never)	0.193	0.174	-0.147 − 0.533	1.112	0.266
Random Effects	**Variance**	**Std Dev**	**Correlation**		
Task Type	0.097	0.311			
Scales' Sub-item ID	0.072	0.269			
Participant ID	1.528	1.236			
Task Type | Participant ID	0.262	0.511	-0.597		
Model Fit	**AIC**	**Log Lik**			
	1009.02	-489.51			
Goal Alignment					
Fixed Effects	**Estimates**	**Std Error**	**95% CI**	**z**	**p-value**
Speech Mode (Non-ISA)	2.309	0.656	1.023 − 3.596	3.518	$<0.001^{***}$
Robot (No)	0.120	0.170	-0.214 − 0.453	0.701	0.483
Voice Assistant (Never)	-0.099	0.316	-0.719 − 0.521	-0.312	0.755
Physical Collaborative Tasks (Never)	0.536	0.270	0.007 − 1.064	1.985	0.047^{*}
Random Effects	**Variance**	**Std Dev**	**Correlation**		
Task Type	0.050	0.224			
Scales' Sub-item ID	0.000	0.000			
Participant ID	4.266	2.065			
Task Type | Participant ID	0.299	0.546	-0.553		
Model Fit	**AIC**	**Log Lik**			
	801.36	-385.68			
Performance Trust					
Fixed Effects	**Estimates**	**Std Error**	**95% CI**	**z**	**p-value**
Speech Mode (Non-ISA)	1.105	0.493	0.138 − 2.072	2.240	0.025^{*}
Robot (No)	-0.041	0.136	-0.307 − 0.226	-0.298	0.766
Voice Assistant (Never)	0.231	0.248	-0.255 − 0.717	0.932	0.351
Physical Collaborative Tasks (Never)	0.400	0.211	-0.014 − 0.814	1.892	0.058
Random Effects	**Variance**	**Std Dev**	**Correlation**		
Task Type	0.015	0.124			
Scales' Sub-item ID	0.225	0.475			
Participant ID	2.630	1.622			
Task Type | Participant ID	0.019	0.136	-1.000		
Model Fit	**AIC**	**Log Lik**			
	990.83	-482.42			
Anthropomorphism					
Fixed Effects	**Estimates**	**Std Error**	**95% CI**	**z**	**p-value**
Speech Mode (Non-ISA)	2.708	0.674	1.387 − 4.03	4.016	$<0.001^{***}$
Robot (No)	-0.168	0.184	-0.528 − 0.192	-0.915	0.360
Voice Assistant (Never)	0.031	0.340	-0.635 − 0.697	0.092	0.927
Physical Collaborative Tasks (Never)	-0.111	0.288	-0.676 − 0.454	-0.385	0.701
Random Effects	**Variance**	**Std Dev**	**Correlation**		
Task Type	0.000	0.000			
Scales' Sub-item ID	0.101	0.318			
Participant ID	5.092	2.257			
Task Type | Participant ID	0.832	0.912	-0.552		
Model Fit	**AIC**	**Log Lik**			
	1239.07	-606.54			