In: *Proceedings of the 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 2023, Busan, South Korea.

Indirect Politeness of Disconfirming Answers to Humans and Robots

Eleonore Lumer*[1], Clara Lachenmaier*[2], Sina Zarrieß[2], Hendrik Buschmeier[1]

***Abstract*— Politeness is a social and linguistic phenomenon that humans use in communication to build and maintain relationships and spare others' feelings. Research on whether humans also apply politeness strategies when interacting with robots – artifacts that lack feelings – yields contradictory findings. This paper presents a human–robot interaction study (N = 40) and compares participants' use of face saving politeness strategies in their responses to disconfirmation eliciting and face-threatening questions asked either by a robot or a human. An analysis of the linguistic properties of participants' answers (response type, use of politeness markers) shows a higher use of indirect politeness in disconfirming answers directed at humans than at robots. This contradicts previous theories on the automatic and 'mindless' application of social strategies towards artificial agents. Alternative explanations for the differences in politeness behavior are discussed.**

I. INTRODUCTION

Politeness is pervasive in human interaction as speakers use it to spare other's feelings, that is, to 'save face' [1], or to form and maintain relationships. Findings and theories on politeness in human–robot interaction are somewhat conflicting. While some studies have argued that users seem to use politeness when interacting with machines even though they are known not to have feelings [2], other research found that users would not use politeness when interacting with robots [3]. This latter conclusion was drawn based on qualitative interviews in which no robot was present – even though the appearance and embodiment of robots have been found to influence user behavior [4], [5]. It is also important to note that the linguistic behavior of users interacting with robots (or conversational agents more generally) could change over time as such systems are increasingly available in the wild and technology is advancing [5]. Research on politeness in human–robot interaction is relevant not only because of the potential benefits of implementing politeness in robots [6], [7], but also because it can provide insights into user behaviour that are interesting from a design perspective. This latter point includes possible insights into whether language choice towards robots is 'mindless' [2], [5], [8].

With the aim to shed more light on the conflicting findings in previous research, this paper presents a human–robot interaction study to individually analyze and compare participants' use of indirect politeness strategies towards a humanoid social robot (the 'Furhat' robotic head; Furhat Robotics, Stockholm, SE) and a human interaction partner. Specifically, the study analyzes linguistic markers for indirect politeness used in disconfirming answers of German speakers. Presenting a novel approach to studying the use of linguistic politeness toward a robot, this paper provides insights into subtle – but relevant – differences in participants' social linguistic strategy in their language use toward robots and humans. To our knowledge, no comparable analysis of linguistic markers of politeness has been carried out. As previous research generally suggests interpersonal differences in the perception of social robots [9], this paper considers interpersonal variations in politeness use between participants as well. The study compares and analyzes indirectness and the use of linguistic markers of politeness in participants' responses to a set of face threatening questions that are posed first by the robot and then, again, by a human interaction partner.

As a concrete example, consider one participant's (no. 33) answer to the question, if they agreed with the robot's assessment of their skills. When the question was asked by the robot, their answer *No* was a clear disconfirmation. When the same question was asked by the human study leader later on, the participant answered (translated) *The not ehm I had the feeling somehow the robot has estimated me lower that I would estimate myself although I already I would say reflects then actually what one can mhm*. This answer differed not only in size, but also in the use of filled pauses (*hem*), hedging (e.g., *somehow*), and in that elaborations are made.

Findings overall show differences in the politeness behavior of participants towards a robot compared to a human in that they used more complex and indirect politeness towards the human. These findings contradict previous research on automatic transfer of social behavior from human interaction to human–robot interaction [8]. Our work thereby contributes to the debate on the use of social behavior of users in human–robot interaction and provides insights into possible alternative explanations for this behavior.

The paper is structured as follows: Section II introduces the concept of linguistic politeness and findings on politeness use in human–robot interaction. Section III motivates the research question and formulates hypotheses. Section IV describes the study design, data collection, and analysis methods. The results of the study are presented in Section V and discussed in Section VI. Section VII concludes the paper.

II. BACKGROUND

A. Linguistic Politeness

There are several ways to define the complex linguistic phenomenon of politeness. According to the most prominent theory [10], politeness strategies are used to save the *face* of the listener or speaker. Face, in this context, is considered to be the public self image of a person [11]. The theory proposes

*These authors contributed equally.

[1]Digital Linguistics Lab, Faculty of Linguistics and Literary Studies, Bielefeld University, Germany. eleonore.lumer@uni-bielefeld.de

[2]Computational Linguistics Group, Faculty of Linguistics and Literary Studies, Bielefeld University, Germany. clara.lachenmaier@uni-bielefeld.de

three top-level influences on politeness strategy choice: the *power* of the listener over the speaker, the *social distance* between the interlocutors, and the *rank of imposition* of an utterance. It should be noted, though, that a number of other influences on politeness have been found, such as gender [12], the presence of a third person, or a person's mood [13].

One linguistic and interactional phenomenon that is of particular interest for politeness research are *disconfirming answers*. These are often face threatening, and interlocutors may need to use them, for example, in response to requests or yes-no-questions. The standalone use of the negative answer particle *nein* (German for 'no') is therefore often omitted in conversations as this could potentially end a conversation [14]. The use of *nein* is not necessarily disconfirming. When confirming a negatively marked statement, *nein* might turn into the answer format of choice. Even in those situations, however, people tend to use face saving variations of *nein* and mitigation techniques such as adding explanations [14]. Only in specific interrogative communication formats, such as performing an anamnesis interview, where brief answers are wanted, standalone *nein* seems acceptable as an answer. In German, one mitigated version of *nein*, which is more accepted and frequent also in standalone answers, is *nee* [14]. Other verbal strategies that mitigate the disconfirmation of *nein* exist as well, such as *hedging* [10] or *expansion sequences* that explain the rejection [14]. Furthermore, *delayed starts* or the use of *hesitation markers* such as *ehm* 'uhm' can also be used to make an utterance more polite [15].

B. Politeness in Human–Robot Interaction

Politeness as a social phenomenon has also been of interest in human–robot interaction research. Analyzing users' preferences regarding the use of politeness by robots, it was found that a general use of politeness in different forms can, for example, improve users' readiness to help [16], compliance [17], and trust [6], as well as the robot's persuasiveness [7]. Contradicting these findings, it was also found that a robot's use of social linguistic strategies such as vagueness or indirectness is often perceived as inappropriate, especially when it does not match other properties of the robot, resulting in a verbal uncanny valley effect [18].

A different line of research investigated whether humans use politeness when interacting with machines more generally. It was found that the normative response bias, which in human interaction leads study participants to respond in a certain way in order to be polite, might be applicable to machines as well [19]. In the study, computers were more positively evaluated when responses were collected with the same computer rather than on paper or using a different computer. Studies such as this helped establish the influential 'Computers Are Social Actors' theory (CASA; [8]), which argues that users mindlessly apply social rules, such as politeness, when interacting with machines [8]. CASA is currently being challenged from different angles: While some research adapts it in order to account for the increased availability of agents and the technological advances that influence their human-likeness [5], others found that communication with agents is not regarded as comparable to human communication as there are vastly different aims, for example the lack of a relationship-building goal [3]. This finding is in line with the idea that the concepts underlying social linguistic strategies, such as the concept of face, might not be transferable to interactions with artificial agents [20]. A recently proposed theory of robots as depictions of social actors [21], tries to account for the discrepancy between users' perception of the robot as a machine without face on the one hand and their use of social strategies, such as politeness, on the other hand.

Given these contradictory findings, this paper studies the use of indirect politeness in human–robot interaction using a novel experimental approach. We study participants' answers directed either at a robot or a human, analyzing their use of answer particles and other linguistic markers for expressing indirect politeness in disconfirming answers.

III. RESEARCH QUESTIONS AND HYPOTHESIS

Based on the research presented above, we formulate the following research questions and expectations. The general research question we pursue is whether human interaction leads to more politeness, e.g., face saving actions, than human–robot interaction. Based on the findings on German disconfirming answers using *nein* [14], we investigate whether participants produce more answers with a 'clear' *nein* to robots than to humans (RQ-1). Furthermore, we analyze whether there is a systematic difference in the realization of disconfirming answers directed at robots and humans (RQ-2). Based on findings of [3], [20], we expected participants to have different concepts of face for the robot and the human, which in turn leads them to choose different degrees of politeness. Specifically, we hypothesize that face threats are less relevant in human–robot interaction than in human interaction (H-1).

Our expectation overall was that participants would try to save face by being indirect when having to answer with a disconfirmation. Specifically, we expected this to manifest in the use of variations of the German answer particles *nein* (such as *nee* or *nö*) that are more face saving [14], as well as in the use of linguistic markers such as hedging [10], additional explanations [14], and hesitation markers [15] accompanying the answer particles.

Additionally, as Brown and Levinson [10] and others [12] suggest that power and distance influence the choice of politeness strategy, we also queried participants' evaluation of their power over the robot and their perceived distance to it. Based on previous research [22], we expect participants to feel generally distanced towards the robot and to not have much power over them.

IV. METHODS

The research goal of this paper is pursued with a novel study design that elicits disconfirming answers by participants using questions with different degrees and directions of face threats. Answers to these questions were either face threatening for the participant giving the answer, or to the interaction partner asking the question (a robot or a human).

TABLE I

FEEDBACK QUESTIONS POSED TO STUDY PARTICIPANTS IN THREE CATEGORIES (A, B, C) BY THE ROBOT AND THE HUMAN STUDY LEADER.

QID	Face threatening for	Question posed by Robot	 Human (researcher)
A1	no one (baseline)	*Do you have any further questions?*	*Do you have any further questions?*
A2		*Have you ever interacted with a robot before?*	*Have you ever interacted with a robot before?*
B1	robot | researcher	*Did you like the test?*	*Did you like the test?*
B2		*Do you think I did a good job with the test?*	*Do you like the design of our test?*
C1	participant	*Was the test difficult for you?*	*Was the answering of the questions difficult for you?*
C2	(& researcher)	*Did you find the assessment of your skills accurate?*	*Did you find Furhat's assessment of your skills accurate?*

A. Study Design

Data was collected in a human–robot interaction study. The study leader introduced the experiment as follows: they designed and programmed a novel German language proficiency test administered by a social robot. The participant's task would be to pilot-test it. The test and the evaluation given to participants were based on a fixed script that was the same for every participant. The actual test performance was disregarded, and participants' language proficiency was always evaluated negatively by the robot in order to elicit face threats. This was done as to control for the face threatening potential of the questions and to make the situation comparable between participants.

Directly after the language test and the performance evaluation, the robot asked participants for feedback using questions that could be answered in confirming or disconfirming ways. Afterward, the human study leader asked the same, or comparable, questions. For the analysis, we chose six questions in three categories of different face threatening potential (see Table I): Two baseline questions that were not face threatening (A1 and A2), two questions that were face threatening to the person (or robot) asking the questions (B1 and B2), and two questions that were face threatening to the participant answering the question (C1 and C2).

B. Data Collection

After signing the informed consent form, the study procedure was explained to participants, and the robot was revealed (Figure 1 illustrates the set-up). The study leader left the room, and the interaction with the robot – controlled, unbeknownst to the participant, from the room next door using a Wizard-of-Oz setup [23] – started. After the interaction, participants were then asked the feedback questions again and filled out a brief questionnaire afterwards. In a debriefing, participants were informed of the actual purpose of the study, the non-informativity of the negative evaluation of their test performance by the robot, and its purpose. The study was approved by Bielefeld University's ethics review committee (application no. 2022-250). The study procedure and the question catalog can be found in the supplementary material: https://doi.org/10.17605/OSF.IO/57X9Y.

C. Participants

Forty German native speakers (23 female, 16 male, 1 non-binary), most of them students (95%) with an average age of 22.9 years ($SD = 4$), were recruited at university campus and offered a compensation of 5 EUR to participate in the study that took approximately 20 minutes. Demographic data, participants' technical knowledge and interest in technology (on five-point scales: *very high, high, average, low, very low*), and their previous experience with robots and voice assistants was collected using a questionnaire at the end of the study.

Fig. 1. Photo showing the study set-up during the language evaluation test and feedback questions by the robot.

While a majority of participants had never interacted with a robot before (72.5%), most had interacted with a voice assistant (82.5%). A majority of participants rated their technical knowledge as average (62.5%), 22.5% as high or very high, and 15% as low or very low. Their interest in technology was rated slightly higher, with 40% rating it as high or very high, 45% as average, and, again, 15% as low or very low.

D. Analysis

Automatic transcripts were generated using the Python package SpeechRecognition (which utilizes the Google Speech Recognition API and, as a backup, CMU Sphinx [24]) and manually corrected by the first two authors. Coding of answers was carried out semi-automatically: features were extracted automatically from the transcripts and manually controlled and supplemented if information was missing.

Three different *types of answers* and three different *types of politeness markers* were distinguished during coding. The type of an answer was coded as either a clear confirmation (+), clear disconfirmation (−), or an indirect disconfirmation (∼). Table II provides an overview with examples of the

TABLE II

EXAMPLES AND EXPLANATION FOR CODING OF THE ANSWER TYPES

Answer type	Short	Explanation of coding
Clear confirmation	+	positive answer particles, such as *ja* 'yes' and variants (*jo*, *joa*), are used
Clear disconfirmation	–	negative answer particles, such as *nein* 'no', and variants (*nee*, *nö*), are used
Indirect disconfirmation	~	(i) no answer particles, but a disconfirmational wording, e.g., using *nicht* 'not' (ii) positive answer particles and a disconfirming explanation or verbal strategy markers (iii) unclear responses that are very indirect and face saving

coding for each *answer type*. Clear confirmations and clear disconfirmations were coded by checking the answer particles (e.g., *ja* 'yes' or *nein*) that were uttered.

Indirect disconfirmations did not include clear answer particles but cover three other ways of expressing disconfirmation, namely (i) negative particles, e.g., *nicht* 'not', (ii) positive answer particles with disconfirming explanations, and (iii) unclear responses. Only very few answers were of this latter type. These were phrased so indirectly that no clear interpretation was possible. For present purposes, these cases can be categorized with other indirect confirmations, as they were also used to save face.

Our analysis mainly considers the two types of disconfirmations, as clear disconfirmations are face threatening and indirect disconfirmations are used to save face. Confirmations, in contrast, cannot be interpreted clearly as face saving, as it cannot be deduced whether participants are uttering a white lie in order to be polite and save face, or telling the truth.

The coding of *politeness markers* was based on markers of verbal strategies, namely the use of hesitation markers (such as *äh* 'uh', *ehm* 'uhm', ... [10]), hedging (such as *vielleicht* 'maybe', *glaub ich* 'I think', ... [14]), and providing an explanation after an answer particle [15]. A detailed example of the coding process, based on the introductory example of participant 33, can be found in the supplementary material.

After coding all interviews, we quantitatively analyzed the data by comparing participants answers depending on whether the question was asked by the robot or human. This is presented in the following section.

V. RESULTS

We start our result section by briefly touching on participants' perceived relationship to the robot, which is relevant for the interpretation and discussion of the result. Following this, we present a mostly descriptive analysis of participants' answers (coded as in Table II) to the feedback questions (Table I). This is followed by an analysis of the politeness markers participants used in their answers.

A. Power and Distance

Participants evaluated their perceived *power* over and *distance* to the robot (using the questions of Lumer and Buschmeier [22]). Participants evaluated their power over the robot to be 48.3 on average (Mdn = 49, SD = 21.2) on a scale from 0 (low authority) to 100 (high authority). Meaning that participants did not perceive to have much power over the robot. Participants evaluated their distance to the robot to be 65.23 on average (Mdn = 66, SD = 21.5) on a scale from 0 (very close) to 100 (very distant). This is a relatively high distance towards the robot when comparing it to the evaluations given by participants in [22].

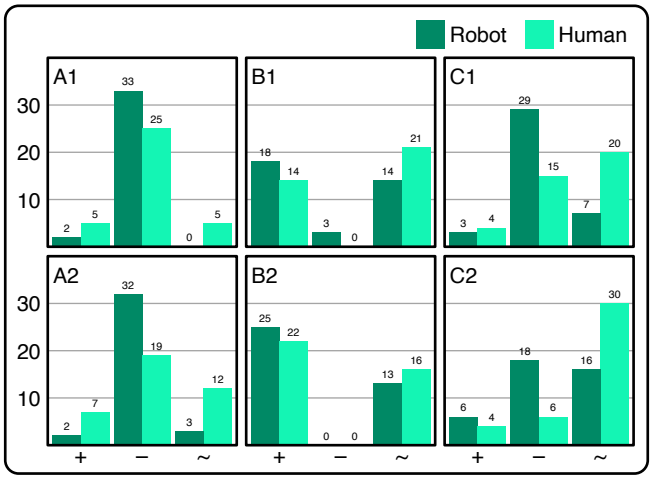

Fig. 2. Answer type distributions for each question by all participants in each condition (human and robot). Each plot shows the three different answer categories that were used: *clear confirmations* (left, +), *clear disconfirmations* (middle, –) and *indirect disconfirmations* (right, ~). The answers given to the Furhat robot are displayed in dark green, while those given to the human interviewer are displayed in light green.

B. Response Types and the Use of Indirect Answers

Figure 2 shows the distribution of answer types (clear confirmation, clear disconfirmation, indirect disconfirmation) for each of the six questions (A1, A2, B1, B2, C1, C2) and for each addressee (robot, human). We observe similarities in answering patterns within question pairs and differences across question pairs and will analyze them individually in the following: We begin with an analysis of participants' answers to the two questions with no face threatening potential (A1 and A2). It can be seen that most participants answered with clear disconfirmations – regardless of whether the question was asked by the robot or the human study leader. Some participants produced indirect disconfirmations, more so when responding to the human.

Next, we analyze how participants responded when their answer poses a face threat to the interaction partner asking the question (B1 and B2). For both questions, participants either answered with a clear confirmation or with an indirect disconfirmation (three participants answered question B1 with a clear disconfirmation to the robot). A small difference in the answers directed at robots and humans can be observed: robots received slightly more clear confirmations, humans more indirect disconfirmations. This difference is not statistically significant though (Fisher's exact test, B1: $p = 0.093$, B2: $p = 0.64$, $\alpha = 0.05/4 = 0.0125$).

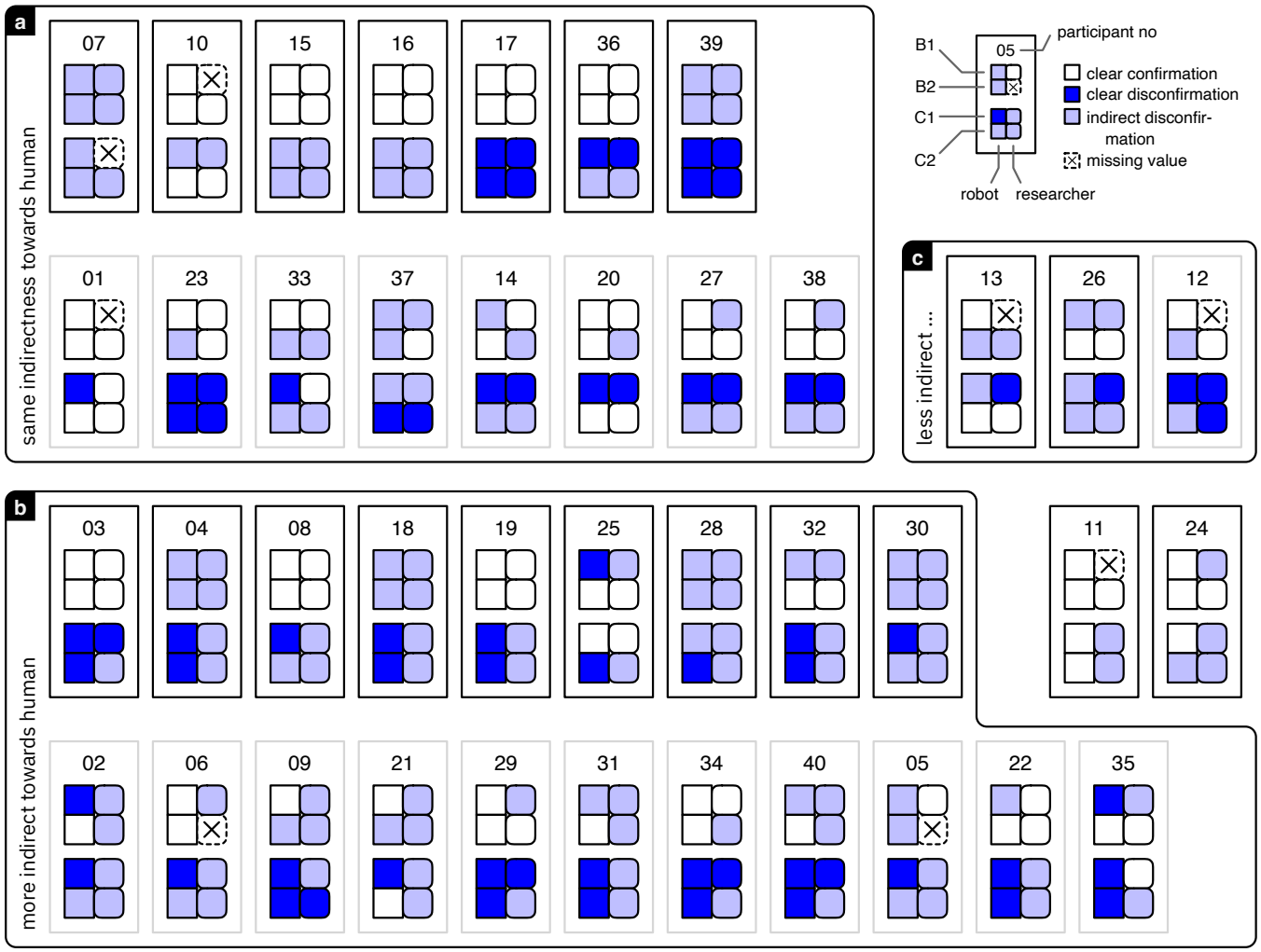

Fig. 3. All 40 participants grouped according to the overall change in indirectness of answers given in response to four questions asked (first) by the robot and (later again) by the human study leader: (a) participants that tend to use the same level of indirectness to the robot and the human, (b) participants that (at least once) gave more indirect answers to the human, (c) participants that (at least once) gave less indirect answers to the human. Each box displays the eight answers that this specific participant gave as small squares. Color reflects the answer type (white: clear confirmation; dark blue: clear disconfirmation; light blue: indirect disconfirmation) and shape reflects the agent who asked the question (square: robot; rounded square: human). The questions are, from top to bottom, B1, B2, and C1, C2. Answers missing in the data are crossed out and not taken into account.

Finally, we analyze how participants responded when their answer to the questions poses a face threat to themselves (questions C1 and C2). When asked by the robot whether the test was difficult for them (C1), a majority of participants (74%) responded with a clear disconfirmation. Responses to the human, in contrast, were more distributed between clear disconfirmations (38%) and indirect disconfirmations (51%). This shift in participants' answers depending on who asked the question – almost half of them first provided a clear disconfirmation to the robot and then responded differently to the human, mostly using an indirect disconfirmation – is statistically significant (Fisher's exact test, $p = 0.003$, $\alpha = 0.0125$). When asked by the robot about its assessment of their skills (C2), participants either responded with a clear disconfirmation (45%) or an indirect disconfirmation (35%). Here, responses to the human were more consistent between participants: a majority produced an indirect disconfirmation (75%). This difference in response to the robot or the human is statistically significant as well (Fisher's exact test $p = 0.0039$, $\alpha = 0.0125$).

C. Individual Differences in Responses

It is important to consider individual differences between participants. Figure 3 visualizes participants' responses to the four non-baseline questions (B1, B2 and C1, C2) asked (first) by the robot and (later again) by the human study leader.

We grouped participants based on their overall response behavior, specifically whether and how their responses to a question changed when asked by the human instead of by the robot: do participants use the same level of indirectness to the human and the robot, or are they more indirect or less indirect to the human than to the robot.

The first group (Figure 3a) shows participants who responded (mostly) in a consistent way, that is they produced a response of the same answer category (and thus level of indirectness) regardless of whether the question was asked

by the robot or the human. Seven participants are clear cases for this category as they always respond in the same way. We also included eight less clear cases (gray outline) in this category who in one question change responses from a clear disconfirmation or an indirect disconfirmation to a clear confirmation, or from a clear confirmation to an indirect disconfirmation. Of the 60 instances of the questions in this group, 48 were responded to in the same way (80%), nine were responded to with a change in answer type (15%). For three question instances, one response is missing.

The second group (Figure 3b) shows participants who responded more indirectly (at least once) when the question was asked by the human. Nine participants are clear cases for this category. We also included eleven participants (gray outline) who in addition to responding more indirectly, produced a response that either changed to or from a clear confirmation. Of the 80 instances of the questions in this group, 41 were responded to with a change in answer type (51%) and 37 were responded to in the same way (46%). For 29 changes, the answer type became more indirect (71%), for twelve the change involved a clear confirmation (29%). For two questions, the instance of one response is missing.

The third group (Figure 3c) shows participants whose responses became less indirect (or did not change) when the question was asked by the human. Two participants are clear cases with a change from an indirect disconfirmation to a clear disconfirmation. One participant (gray outline) additionally changes an indirect disconfirmation to a clear confirmation. Two participants (11, 24) did not fit into the three groups.

When looking at question C1 as a concrete example, 13 participants (33%) responded with a clear disconfirmation to both the human and the robot, and the same number of participants changed their response from a clear disconfirmation towards the robot to an indirect disconfirmation when answering to the human.

The overall tendencies to change the answer type between conditions is confirming the results described in the previous subsection and shown in Figure 2.

D. Politeness Markers: Human–Robot vs. Human Interaction

When analyzing the politeness markers that participants used in their answers, clear differences can be observed between answers directed at the robot and answers directed at the human. To exemplify this, Figure 4 displays the difference in usage of 'hedging' for all six questions. As can be seen, participants barely used hedging when interacting with the robot, while it was consistently used more in interaction with the human. As mentioned above, we also compared the use of other politeness markers: hesitations (such as *äh/eh* 'uh', *ähm/ehm* 'uhm') as well as further talk after the answer particle or answer including explanations (graphs can be found in the supplementary material). Overall, all politeness markers were consistently used more in human–human interaction than in human–robot interaction.

VI. DISCUSSION

The results concerning participants' perception of power and distance towards the robot are in line with our previous study of robot perception [22]. Generally, the relationship between the user and the robot is evaluated to be rather distant in this experiment. The power over the robot was not perceived to be particularly high, which could result from a perceived lack of control over the robot in the study set-up. This is also in line with our previous evaluations made by participants from a fictional third-person perspective where the power over the robot was evaluated to be rather small when it was in a public space [22]. Politeness theory [10] suggests that these perceptions of high distance and low power over the robot should lead to the use of face saving politeness strategies towards the robot (as the potential face threat for the robot is high). In the following, we analyze whether that was the case in the interaction with the robot in our study.

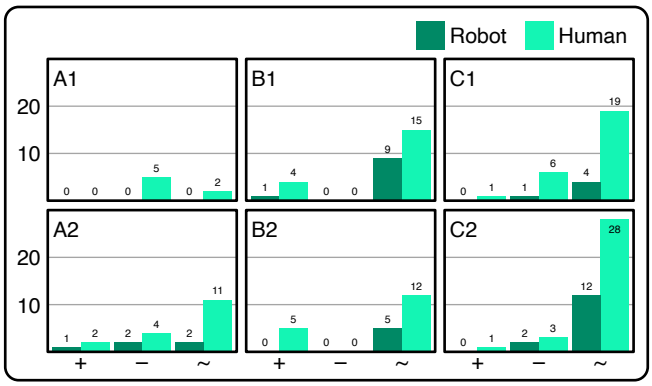

Fig. 4. Hedging behavior by participants for each condition and in each question. Each question plot displays the three different answer categories: *clear confirmations* (+), *clear disconfirmations* (−), and *indirect disconfirmations* (~). The answers towards the human are displayed in light green, while those to the robot are displayed in dark green.

As can be seen in Figure 2, the responses to the two baseline questions (A1 and A2) clearly differ from the answering behavior to the four face threatening questions. This can be seen as a validation of our procedure: without face threat, most participants do not hesitate to use clear disconfirmations – in contrast to the face threatening questions.

When analyzing the differences between participants' answer types for each question based on Figure 3, we found individual differences between the answering behaviors towards humans and robots. Some participants responded with the same answer type to the robot and the human, while most, at least for some questions, changed their answer depending on the interlocutor. This is in line with previous research on individual differences in user behavior towards robots [9]. Already with this display of the answer type results in Figure 3, the tendency of most participants to use more indirect and hence polite disconfirmations towards the human can be observed.

Overall, when analyzing the answer type provided by participants to the different questions based on Figure 2, we found that some users, when wanting to disconfirm, used an indirect disconfirmation also in interaction with the robot. This was the case for the questions that were face threatening for the interaction partner asking the question, robot or human

(B1 and B2). This can be seen as a use of indirect politeness in interaction with the robot by some participants. However, more indirect disconfirmations were used towards the human in these questions as well. When comparing this to the questions that were also face threatening to the participant (C1 and C2), we found a clear difference in polite language use in answers directed at robots and humans. As for both questions, the answers towards the human were overall more indirect than towards the robot.

Our findings for the politeness markers clearly show more use of these markers in human interaction than in interaction with the robot. This again shows the difference in politeness behavior of users towards the robot and the human. Based on previous research, these markers are used to diminish the face threat and, therefore, as politeness strategies when added to a disconfirmation [10], [14], [15].

We argue that these differences do not result from mirroring or alignment of politeness behavior. This phenomenon occurs when a person adapts to their interlocutor's use of politeness [25], [26]. Mirroring can be excluded in this case, as the researcher read out loud previously formulated questions (B1, B2, C1, C2; see Table I and supplementary material for original German wording), thereby avoiding hedging. This additionally also assured the regularity of questions between participants. Since neither the human nor the robot hedged when asking the question, differences in participants' use of hedging shown in Figure 4 cannot be attributed to mirroring, providing further evidence for the different politeness behavior exhibited in human–robot interaction. Hence, next to the actual response type, the increased use of politeness markers with humans shows an overall more complex use of politeness with a human than with a robot. Overall, this shows a clear difference in politeness behavior and hence face consideration between human and human–robot interaction.

This stands in opposition to the CASA theory [8]. Our data shows this, as even though for some questions that are face threatening to the person or robot asking the question there is a similar decision on answering type in disconfirmation, the additional use of politeness markers and hence the overall politeness behavior clearly differs between human–robot interaction and human interaction. For the questions also threatening participants' face even more so, as here the difference was already apparent in the answer types, with more indirect answers towards the human.

We overall can therefore accept our H-1 Hypothesis as our data shows a clear difference in politeness behavior and hence also of face relevance between human–robot interaction and human interaction. The different behavior regarding face relevance can be observed with this data. The actual reasons for this behavior can, however, not be clearly derived from a different face perception.

We see two possible explanations for the differences in politeness behavior towards robots and humans shown in our results. First, we believe the differences might result from the perceived abilities of the robot and participants' assumption that the robot is not able to process complex information. This is based on two observations in our data. Some participants changed their response from a clear disconfirmation towards the robot to an indirect disconfirmation towards the human. Further, participants used politeness markers in interaction with the human, however rarely with the robot. Both these observations can be seen as suggesting that participants did not regard the robot or the technology behind it as capable of processing these more complex linguistic strategies.

A second possible explanation is that these results can also be taken to indicate a lack or difference in face perception of the robot. This is the case as the behavior towards the robot can be interpreted as less polite than with the human because they were more direct, and little politeness markers were used in interaction with the robot. Based on the theory of face [10], [11] and the evaluated perception of participants' relation to the robot based on the perceived power and distance, they should however use face saving politeness strategies. This can be seen as in line with the research by Clark [20] and their suggestion that the face concept is not directly transferable to the interaction with artificial agents.

For future research using this approach, we alternatively suggest the use of a between-subject design approach to circumvent possible priming effect discussions arising due to the within-subject design chosen for this study. The design chosen in this experiment was necessary in order to control for individual differences in answering behavior, as suggested by Fischer [9]. Further, we exclude the possibility of priming effects in our results, as our data did not include hints for priming effects. We did, for example, not find a pattern of choosing more positive confirmatory and not face threatening answers in the HHI condition, which could be expected as a result of being primed by already having answered the same question to the robot.

VII. CONCLUSIONS

This paper presents a new methodology to linguistically analyze the use of politeness by users interacting with robots in comparison to humans. The study compares responses uttered to a robot with those uttered to a human. In the analysis, we considered the face threatening potential of the questions posed in order to elicit disconfirmation. The analysis concerned the type of the answer (confirmation, disconfirmation or indirect disconfirmation) as well as politeness markers occurring in disconfirming answers (hedging [10], hesitation markers [15], and additional explanations [14]).

For questions concerning the face of the listener, even towards the robot, when wanting to express disconfirmations, participants rather used an indirect word choice. For the questions also concerning participants' face threats, the results showed a clear difference in polite word choice between human–human and human–robot interaction. Towards the human, the responses were always more indirect than towards the robot. Overall, our data shows more use of indirect responses towards the human than towards the robot.

This is in line with our analysis of politeness markers. Where we observed that almost no markers were used when interacting with the robot, while in human interaction, these were used frequently, especially in disconfirming answers.

Overall, our study shows clear differences in politeness behavior towards a robot and towards a human. Our findings, therefore, contradict the claims of CASA [8], because they show that humans do not automatically answer in the same way to humans as to robots. Concretely, overall politeness behavior differed in the two conditions, especially for the verbal strategy markers. Further, the results can also be seen as contradicting the adapted CASA version by [5]. This is the case as there were individual differences in participants' responses to the robot, as already suggested by Fischer [9].

We make two interpretation suggestions for these results. Either they can be taken to indicate a lack or difference in the face perception of the robot as would be in line with research by Clark [20], suggesting that the concept of face is not directly transferable to interaction with artificial agents. On the other hand, these results can also be taken to indicate a lack of confidence in the interpretation abilities and technical capabilities of the robot. The latter could also be considered to be in line with the view of robots as representations of social actors [21]. This theory would also account for the found individual differences in politeness choice by participants [9].

References

[1] P. Brown, "Politeness and Language," in *International Encyclopedia of the Social & Behavioral Sciences*. Elsevier, 2015, pp. 326–330.

[2] B. Reeves and C. Nass, *The Media Equation*. Stanford, CA, USA: CSLI Publications, 1996.

[3] L. Clark, N. Pantidi, O. Cooney, P. Doyle, D. Garaialde, J. Edwards, B. Spillane, E. Gilmartin, C. Murad, C. Munteanu, V. Wade, and B. R. Cowan, "What makes a good conversation?: Challenges in designing truly conversational agents," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, Glasgow, UK, 2019, pp. 1–12.

[4] A. M. Rosenthal-von der Pütten and N. C. Krämer, "Individuals' evaluations of and attitudes towards potentially uncanny robots," *International Journal of Social Robotics*, vol. 7, pp. 799–824, 2015.

[5] A. Gambino, J. Fox, and R. A. Ratan, "Building a stronger CASA: Extending the computers are social actors paradigm," *Human-Machine Communication*, vol. 1, pp. 71–85, 2020.

[6] S. Kumar, E. Itzhak, Y. Edan, G. Nimrod, V. Sarne-Fleischmann, and N. Tractinsky, "Politeness in human–robot interaction: A multi-experiment study with non-humanoid robots," *International Journal of Social Robotics*, vol. 14, no. 8, pp. 1805–1820, 2022.

[7] A. Iop, "Assessing perceived politeness in a virtual agent's request to join a conversational group," Master's thesis, School of Electrical Engineering and Computer Science, KTH, Stockholm, Sweden, 2022.

[8] C. Nass and Y. Moon, "Machines and mindlessness: Social responses to computers," *Journal of Social Issues*, vol. 56, no. 1, pp. 81–103, 2000.

[9] K. Fischer, "Interpersonal variation in understanding robots as social actors," in *Proceedings of the 6th International Conference on Human-Robot Interaction*, Lausanne, Switzerland, 2011, p. 53–60.

[10] P. Brown and S. C. Levinson, *Politeness: Some Universals in Language Usage*. Cambridge, UK: Cambridge University Press, 1987.

[11] E. Goffman, "On face-work: An analysis of ritual elements in social interactions," *Psychiatry*, vol. 18, no. 3, pp. 213–231, 1955.

[12] T. Holtgraves and J.-F. Bonnefon, "Experimental approaches to linguistic (im)politeness." in *The Palgrave Handbook of Linguistic (Im)Politeness*. London, UK: Palgrave Macmillan, 2017, pp. 381–401.

[13] N. Vergis and M. Terkourafi, "The role of the speaker's emotional state in im/politeness assessments," *Journal of Language and Social Psychology*, vol. 34, pp. 316–342, 2015.

[14] W. Imo, "Über nein," *Zeitschrift für germanistische Linguistik*, vol. 45, no. 1, pp. 40–72, 2017.

[15] A.-B. Stenström, "Pauses and hesitations," in *Pragmatics of Society*, G. Andersen and K. Aijmer, Eds. De Gruyter, 2011, pp. 537–568.

[16] V. Srinivasan and L. Takayama, "Help me please: Robot politeness strategies for soliciting help from humans," *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 4945–4955, 2016.

[17] N. Lee, J. Kim, E. Kim, and O. Kwon, "The influence of politeness behavior on user compliance with social robots in a healthcare service setting," *International Journal of Social Robotics*, vol. 9, pp. 727–743, 2017.

[18] L. Clark, A. Ofemile, and B. R. Cowan, "Exploring verbal uncanny valley effects with vague language in computer speech," in *Voice Attractiveness: Studies on Sexy, Likable, and Charismatic Speakers*, B. Weiss, J. Trouvain, M. Barkat-Defradas, and J. J. Ohala, Eds. Singapore: Springer, 2021, pp. 317–330.

[19] C. Nass, Y. Moon, and P. Carney, "Are people polite to computers? responses to computer-based interviewing systems," *Journal of Applied Social Psychology*, vol. 29, no. 5, pp. 1093–1109, 1999.

[20] L. Clark, "Social boundaries of appropriate speech in HCI: A politeness perspective," in *Proceedings of the 32nd International BCS Human Computer Interaction Conference*, Belfast, UK, 2018, p. 5.

[21] H. H. Clark and K. Fischer, "Social robots as depictions of social agents," *Behavioral and Brain Sciences*, vol. 46, p. e21, 2023.

[22] E. Lumer and H. Buschmeier, "Perception of power and distance in human-human and human-robot role-based relations," in *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction*, 2022, p. 895–899.

[23] N. M. Fraser and G. Gilbert, "Simulating speech systems," *Computer Speech & Language*, vol. 5, pp. 81–99, 1991.

[24] A. Zhang. (2017) SpeechRecognition (Python package, v3.8). [Online]. Available: https://pypi.org/project/SpeechRecognition/

[25] T. Gretenkort and K. Tylén, "The dynamics of politeness: An experimental account," *Journal of Pragmatics*, vol. 185, pp. 118–130, 2021.

[26] M. A. de Jong, M. Theune, and D. Hofs, "Politeness and alignment in dialogues with a virtual guide," in *Proceedings of the 7th International Conference on Autonomous Agents and Multiagent Systems*, Estoril, Portugal, 2008, pp. 207–214.