The effects of a social robot's message design logic and gender in a regulative social context

Utrecht University

Anouk van Santen, 6784976

Department of Information and Computing Sciences, Utrecht University

Master thesis Artificial Intelligence

Supervisor & first examiner: Maartje de Graaf

Second examiner: Susanne Poeller

06-02-2024

Abstract

This study investigates the effects of message design logic (MDL) and gender presentation of social robots on human perceptions in a regulative social context. Building on the work of Edwards et al. (2019a), which explored the impact of different MDLs on user perceptions, this research introduces the variable of the robot's gender. There is not much known yet about how the gendering of social robots affects user's perceptions and interactions. A sample of 162 EU participants evaluated the robot's communicative success, warmth, competence, norm violations, and behavioral tendencies based on three MDLs (expressive, conventional, rhetorical) and two gender presentations (masculine, feminine). The findings aim to provide insights into the preferences for gendered social robots in leadership roles and the influence of gender stereotypes on human-robot interaction. The analysis showed that the MDL affects the user's perception of the robot, but contrary to expectations, gender didn't affect the user's perception. Therefore, it is concluded that it is more important which message the robot conveys rather than the gender of the robot that conveys the message. This research contributes to the design of social robots by highlighting how the gender of a robot and the MDL it uses can affect user acceptance and the perceived effectiveness of robotic communication.

Introduction

There is not much known about the social influences of robots or potential influences of robot in everyday communication. Nevertheless, people are increasingly interacting with AI systems on a daily basis (Powers & Kiesler, 2006; Riedl, 2019). Understanding what this technology means, beyond the technical issues, is vital. This challenge requires an interdisciplinary approach (Shank et al., 2019). Human perceptions and behavioral responses are crucial to gain insights in the potential interactions with robots.

Human-robot interaction (HRI) is a multidisciplinary field considering the human perception of robots and their relationship with robots. It brings together ideas from a wide range of disciplines: engineering, psychology, designing, anthropology, sociology, and philosophy (Bartneck et al., 2020). One topic in this field is social robots, which are designed to evoke meaningful social interactions (Edwards et al., 2019a). Previous work in HRI demonstrated that social robots influence user attitudes and behaviors (Eyssel & Hegel, 2012; Edwards et al., 2019a; Eyssel & Kuchenbrandt, 2011; Jackson et al., 2020; Kuchenbrandt et al., 2014; Perugia & Lisy, 2023). Research in this field is important so HRI designers can utilize social robot design and behavioral cues, including manipulation of the robot's gender presentation, to maximize its persuasive effects and acceptability.

This study aims to partially replicate the study by Edwards et al. (2019a), which focused on the content and goal of a message of a social robot. In this study, a social robot is the leader of a group of students that needs to complete a class project. The robot has to deal with a student who is not doing its job. Edwards et al. (2019a) have already researched the effects of robot's messages with varying levels of sophistication and task-completion goals (message design logics) on the perception of the robot.

The current research adds the manipulation of the gender of the robot. There is not much known yet about how the gendering of social robots affects user's perceptions and interactions (Perugia & Lisy, 2023). This study will investigate user's perceptions about the gendered social robot and the different types of messages it conveys to gain more insight into the preferences regarding the gender of a social robot in a leadership position. The insights can be used for a greater understanding of the human-robot interaction process.

Outline

The next section discusses the theoretical background and previous research on this topic. First, the process of social categorization is explained and its applications are discussed. Social categorization is fundamental for human cognition, and research shows that it is also applied to robots (Eyssel & Kuchenbrandt, 2011). Then, the specific social categorization of gender and the consequences are discussed in the section on gender stereotyping. Stereotyping is an automatic process that guides human behavior (Cislaghi & Heise, 2019). Based on the stereotypes and social norms, the expectations of gendered robots the violations of the expectations are addressed. Lastly, the literature review delves deeper into the speech of a social robot: how communication affects people's attitudes and perceptions of the robot, which is already studied by Edwards et al. (2019).

The methodology section outlines the participants involved in this study, followed by a detailed description of this study's procedure. It is explained how the independent variables are manipulated and how the data is be collected.

In the results section, there is a detailed description of the statistical analyses. Multiple MANOVAs and an ANOVA are performed and the statistically significant as well as the insignificant results are described.

Then it will be discussed what the results mean and what the implications are. The limitations of this study are discussed and there are some proposals for future studies.

In the conclusion, there is a summary of the most important theories, results, and implications of this study.

Literature review

Social categorization

Categorization is fundamental for human cognition (Bodenhausen et al., 2012). The capacity for generalization and inductive reasoning allows for reasoning about something new based on experience or knowledge about the category (Rhodes & Baron, 2019). This applies to several domains of experience. For example, it is possible to speak about the category dogs, without naming every single dog. The experience that one had with a dog can influence the perception of dogs in general for that person. Humans also classify persons into a social category. It guides their behavior, perception, and reasoning about the thoughts and beliefs of other people based on their group membership (Liberman et al., 2017). Social categorization is an efficient process that organizes and structures the knowledge about human attributes and allows for navigating through the complex social environment (Bodenhausen et al., 2012). The categorization can be based on attributes such as age, gender, or ethnicity and the process is nearly automatic (Devine, 1989). The selection of the most relevant category varies according to the context, behavior & characteristics of the target, and the state of the perceiver (Bodenhausen et al., 2012). The process of categorization gives people relevant knowledge that allows for useful inferences for determining whether and how to interact with other people.

Although the process of social categorization is efficient, it can have negative consequences. Simplifying the social environment can lead to disrespectful or negative stereotyping, prejudices, and discrimination. As social categorization is a nearly automatic process, it might be impossible to not have stereotypes or biases. Therefore, people should be aware of their categorizations and minimize discriminatory or harmful behavior.

One of the attributes that humans categorize is gender. It organizes the information about human bodies. For the current study, gender is not only a fixed category, but a system of thought and practice embedded in social structures, behavior, design, and norms (Perugia & Lisy, 2023). Even with the smallest piece of information, people tend to attribute a gender to a humanoid robot. Eyssel & Hegel (2012) found that the visual cue of hair length is enough to attribute a gender to the robot: short hair on the robot was perceived as more masculine than long hair. Other factors that are cues for the gender of a robot are names and pronouns, facial features such as definition or color of lips, voice, clothes in combination with the color of it and body shape such as shoulder width (Perugia & Lisy, 2023).

Previous research shows that humans also apply social categorization to robots (Carpinella et al., 2017; Eyssel & Hegel, 2012; Eyssel & Kuchenbrandt, 2011). Eyssel & Kuchenbrandt (2011) used the social category group membership. They manipulated the label of a robot, making it an in-group or out-group robot, which led to a difference in the human perception of the robot. An in-group robot was rated more favorable, and people anthropomorphized it more strongly than the out-group robot. This in-group bias can be interpreted as a result of social categorization applied to non-human agents.

Gender stereotyping

The process of social categorization activates stereotypes associated with the category being used. Some examples of these stereotypes are: girls are more emotional and boys don't cry (*Effects Of Stereotypes On Personal Development,* 2023). The content of these stereotypes can be highly diverse. According to the stereotype content model (SCM), proposed by Fiske et al. (2002), there are two fundamental dimensions in each stereotype: warmth and competence. Warmth relates to whether others have positive or negative intentions. Competence relates to the capability of others to effectively reach a goal. There is evidence that warmth is the primary dimension and is judged before competence (Fiske et al., 2007). Therefore, warmth judgements have more influence on behavioral reactions.

The stereotypical pattern found for the two genders is that women are perceived as higher in warmth than men and men are perceived as higher in competence than women (Fiske et al., 2002). These overall stereotypes are confirmed in a study by Ebert et al. (2014). Furthermore, in this last study, it is found that there is an association between people's gender and competence. Both women and men associate their gender with competence, but men demonstrated a stronger association.. Based on these findings for human stereotypes, Carpinella et al. (2014) showed that humanoid feminine robots are perceived as higher in warmth than humanoid masculine robots. Contrary to the expectations, the experiment showed that the masculine robots were not perceived as higher in competence. However, the stimuli in this study was only an image of the robot, so results might be different when there is a video of the robot applied in a social domain.

The Behaviors From Intergroup Affect and stereotypes (BIAS) map supposes that stereotypes are linked with behavioral aspects (Cuddy et al., 2007). Warmth stereotypes elicit active facilitation, where people aim to benefit a group and are willing to help or assist others while weakening active harm, the tendency to hurt a group and its interests. Competence stereotypes, on the other hand, elicit passive facilitation, where one accepts obligatory association with a group while weakening passive harm, the tendency to distance a group by diminishing their social worth through excluding, ignoring or neglecting (Cuddy et al., 2007). Although this framework focusses on intergroup relations, Mieczkowski et al. (2019) showed that it also applies to social robots. People ascribe impressions of warmth and competence to robots and these predict the behavioral tendencies, which do not differ from the impressions people have from other humans (Mieczkowski et al., 2019).

Gender stereotypes influence the expectations humans have from another human or robot for a specific task (Winkle et al., 2022). There are some tasks considered typically female and tasks considered typically male. Kuchenbrandt et al. (2014) found in their study that even the smallest detail can determine the gender typicality of a task. The task of sorting a sewing box is perceived as typically female and sorting a toolkit box is perceived as typically male. As a result, the gender typicality of a task influences how successfully participants interact with the robot to complete the task. The leadership position of the robot in the current study might be seen as a typical male task and can affect the perceptions of the robot.

Expectations, norms & norm violations of social robots

When social robots are used in communicative contexts, people have preexisting expectations of the robot's behavior (Levine et al., 2000). These expectations can arise in the process of social categorization and gender stereotyping, but can also be based on social norms. Social norms are unwritten rules of what is considered normal and acceptable in a society and it guides behavior (Cislaghi & Heise, 2019). Social norms can vary for individuals based on factors such as age and gender. For instance, the norms for adults differ from those for children. Similarly, there are distinct norms for different genders; men are often expected to be leaders while women are expected to be followers. This disparity can hinder women from successfully fulfilling leadership positions (United Way of the National Capital Area, 2023).

The differences in norms and expectations are found for robots as well. An experiment from Jackson et al. (2020) found that it is more favorable for masculine robots to reject a command than for feminine robots to do so. This has to do with the gender norms that exist in society where women are expected to be nice and kind and men are more harsh. So, the feminine robot's behavior deviates from what it is expected to do based on gender norms, and this expectation violation results in a preference for a masculine robot in this situation. Furthermore, Jackson et al. (2020) found that robots rejecting a command are perceived more favorable if their gender matches the gender of the command giver and the gender of the participant rating the situation. As mentioned before, humans consider their own status when categorizing other humans. The study by Jackson et al. (2020) demonstrates that humans also consider their own status (gender in this case) when categorizing and evaluating a robot.

Human interactions

Humans are social species in which communication is an important, scripted, or automatic, process (Spence et al., 2014; Kellerman, 1992). The current technology enables humans to communicate with another partner than a human, namely a social robot. This development has led to robots being a participant in communication, where people talk to and with the robot (Guzman, 2018). According to the Computers As Social Actors (CASA) paradigm, people tend to respond to a computer as if it were a person (Reeves & Nass, 1996). Furthermore, Edwards et al. (2019b) showed in their research that the humanoid robot Pepper is sufficiently human-like, such that people apply the same communication tools as they do when communicating with another human. Therefore, human interaction theories are relevant to investigate the communication with robots.

People apply interpersonal and relational norms to interactions with robots (Kim et al., 2019). In some situations, it might be more exciting to have a robot that has novel messages all the time, whereas other situations require the robot to have more predictability instead of novelty. Also, there are some general expectations that people have from other people's communication, known as Grice's maxims (Grice, 1975). The maxims describe that humans expect people to tell the truth; provide the right amount of information; be relevant and strive to be clear. These principles can be applied in human-robot interaction, to meet the expectations of humans in their interaction with a robot.

The way a social robot communicates with humans is likely to affect people's perceptions of the robot's effectiveness and appropriateness (Edwards et al., 2019a). Message

Design Logic (MDL) is a theoretical framework, described by O'Keefe (1988) as "The kind of communication-constituting belief system the message producer relies on in reasoning from the goals sought to the message design used". So, there must be a logic in the message that has a connection to the goal that is being pursued. Due to individual differences in theories of communication, there are variations in message production and interpretation processes (Edwards et al., 2019a).

According to the framework of O'Keefe (1988) there are three different MDLs which vary in the level of sophistication. The least sophisticated MDL is the expressive logic where communication is a mean for conveying thoughts and feelings of the speaker and the message does not necessarily help with achieving a goal. The conventional MDL is somewhat sophisticated and is viewed as communication according to socially conventional rules and procedures. Rather than straightforwardly sharing thoughts and feelings, a conventional message uses the existing structure of rights and obligations in order to achieve the persuaded goal (Edwards et al., 2019a). Lastly, the rhetorical MDL is the most sophisticated. Rather than accepting the conventional social rules, a rhetorical message is to reshape situations and identities to create a social context in which it is the most likely to reach a goal (Caughlin, 2010).

The MDLs are arranged hierarchically in terms of functionality in complex interactions. The expressive MDL is the least competent type of communication, rhetorical is the most competent type of communication and conventional is in between (Edwards et al., 2019a). Furthermore, O'Keefe (1988) argues that there is an ordering in the MDLs based on developmental priority. One must first be able to express thoughts and feelings (expressive MDL) before these expressions can be used to achieve a goal (conventional MDL). This has to be mastered and normative definitions of situations and identities must be learned before it can be reshaped (rhetorical MDL).

The study by Edwards et al. (2019a) found that the robot using the rhetorical MDL was perceived as most favorable on many variables (e.g. face support, credibility, motivational effectiveness). The current study adds the manipulation of the gender of the robot to the study by Edwards et al. (2019a). Based on the findings of the previous study and theoretical background discussed above it is expected to find the same results as Edwards et al. (2019a) regarding the MDLs. However, the addition of gender might influence the effects, as there are different expectations from the genders. Women are typically seen as more caring and therefore the rhetorical message might be rated better for feminine robots than for masculine robots. Furthermore, the role of leadership might not be seen as typically feminine, so feminine robots can be perceived as less competent for completing the task.

H1: Rhetorical messages are perceived as the most favorable in perceptions of communicative success and goal-relevant attributes, followed by conventional and expressive messages respectively.

H2: The expressive MDL is rated more favorably for masculine robots than for feminine robots (in perceptions of communicative success and goal-relevant attributes), as it shows less compassion and executes the team member from the job.

H3: Social robots using rhetorical MDL in their speech are evaluated the most favorable on the dimensions of warmth and competence.

H4: Masculine robots are perceived as more competent than feminine robots in the role of leader and feminine robots are perceived as higher in warmth than masculine robots.

H5: Rhetorical messages score the highest on the behavioral tendencies associated with warmth and competence: active facilitation and passive facilitation respectively.

H6: Masculine robots score higher than feminine robots on the BIAS map for active facilitation and lower than feminine robots for passive harm.
Feminine robots score higher than masculine robots on the BIAS map for passive facilitation and lower than masculine robots for active harm.

H7: Feminine robots violate the norm more than masculine robots when using the expressive MDL.

Methods

This study has a quantitative experimental research design and the following research question is addressed: What influence does the MDL (expressive, conventional, rhetorical) and gender (masculine or feminine) of a social robot in a regulative context have on people's ratings of the message (perceptions of communicative success, goal-relevant attributes) and their evaluation of the robot (perceptions of its warmth, competence & discomfort, norm violations, and behavioral tendencies)?

Participants

The sample included 162 EU participants (86 women, 76 men, and 1 non-binary), recruited and compensated through Prolific. Their ages ranged from 20 to 70, with a mean of 37.74 (*SD* = 13.03). The participants self-reported that they have not encountered a lot of robots in the last year, with a mean of 1.94 on a 7-point Likert scale (*SD* = 1.02). The mean of their knowledge about robots and/or robotic domain is 2.41 on a 7-point Likert scale (*SD* = 1.14).

Procedure

To answer the research question, a quantitative, experimental, between-subjects research design is be used. This study has a 2 (gender) x 3 (MDL) design. Participants are randomly assigned to the gender of the robot and the MDL and equally distributed across the six conditions. Participants filled in an online questionnaire, created with Qualtrics. The questionnaire includes questions about demographics (age, gender, and academic background). The robot Pepper will be masculine or feminine, which is made clear by the use of a name, corresponding pronouns, and the sound of the voice. At the start, the participants watch a video of the robot Pepper, telling the following scenario:

> Hello. My name is [Alexander/Alexandra], an autonomous social robot created to facilitate projects in educational environments. A course instructor assigned me to supervise a small group of college students tasked to complete a class project. My performance will be evaluated based on the group project's outcome. Each student will

> receive both an overall grade for the quality of the project report and an individual grade based on their contribution to the group effort. It is my responsibility as a supervisor to communicate these grades to the instructor. Among the students is Ron, who has been causing some problems. He frequently arrives late to group meetings and entirely skipped one meeting without prior notice. This prompted discussions within the group about his potential removal. Ron attended the subsequent meeting, expressed regret for missing the previous meeting, and attributed it to family problems. He then offered to undertake research on one important aspect of the project, claiming a special interest in that topic. With the project due next week, the group planned to finalize the report during our meeting, scheduled for tomorrow afternoon. Ron contacted me today and informed me he has not completed his research yet and requires more time.

This scenario is slightly adapted from Edwards et al. (2019a), such that it can be used as a first-person story from Pepper. After the video, there is a confirmation check to confirm that the participant understood the main essence of the scenario. Then the participants watch another video, in which Pepper conveys a message to Ron about his behavior in the group. The voice of Pepper matches its gender. Each participant will see one video, in which Pepper conveys a message corresponding to one of the MDLs (see Appendix A for full messages). After this video, there was again a confirmation check to confirm that they understood the main essence of the message. Then, the participant answers a set of questions, to measure the dependent variables.

Measures

All the following questions are randomized for each participant. Participants will answer some questions about the performance of the robot in this specific scenario, which were used in the study by Edwards et al. (2019a). On a 7-point Likert scale (1 = extremely unlikely, 7 = extremely likely), participants rate the likelihood that the message target Ron will complete the task; that the group will complete the task; that Ron is satisfied with Pepper's performance as a leader and that the group is satisfied with Pepper's performance as a leader ($\alpha = 0.74$). On another 7-point Likert scale (1 = extremely ineffective, 7 = extremely effective), participants rate the effectiveness of Pepper's message in motivating the target message Ron, in showing consideration, understanding, and empathy ($\alpha = 0.94$).

Participants will also answer the question "To what extent does the robot's behavior go against the norms of the society?" on a 7-point Likert scale, to measure norm violations (Bräuer & Chaurand, 2009).

Furthermore, the participants complete the Robot Social Attribute Scale (RoSAS). The three dimensions of this measurement scale (warmth ($\alpha = 0.850$), competence ($\alpha = 0.92$) and discomfort ($\alpha = 0.81$)) are proven to measure independent constructs of robot perception (Carpinella et al., 2017). Participants will answer the question "Using the scale provided, how closely are the words below associated with the robot Pepper you just saw in the video?" for 18 items on a 7-point Likert scale (1 = definitely not associated, 7 = definitely associated).

Lastly, participants fill in the Behaviors From Intergroup Affect and Stereotypes (BIAS) map (Cuddy et al., 2017). For the BIAS map, participants rate the likeliness of behavioral tendencies a 7-point Likert scale (1 = very unlikely, 7 = very likely) on four 3-item scales: active facilitation (assist, help, protect, $\alpha = 0.87$), active harm (attack, fight, harass, $\alpha =$

0.82), passive facilitation (associative with, cooperate with, unite with, α = 0.80) and passive harm (exclude, ignore, neglect, α = 0.82). See Appendix B for full questionnaire.

Stimuli

Gender

To verify the effectiveness of the gender manipulation, 34 participants listened to four audio fragments. Two masculine robot voices (Alex and Evan) and two feminine robot voices (Samantha and Allison), from the Apple system voices, uttered the neutral statement: "According to my watch, it is now a quarter to three. The train will leave in five minutes." (adapted from Kuchenbrandt et al., 2014). Then the participants answered the following questions for each voice on a 7-point Likert scale (1= not at all, 7= extremely): (1) To what extent do you perceive the voice as masculine? (2) To what extent do you perceive the voice as feminine? (3) To what extent do you perceive the voice as androgynous? (4) To what extent do you perceive the voice as robot-like?

The statistical analyses (MANOVA) show that the voices were perceived as significantly different (Pillai's trace = 0.79, F(3, 12) = 2.70, *p* = .004). Corresponding ANVOVAs and post hoc tests revealed a significant difference on the masculine scale (F(3,12) = 22.38, p < .001) between the masculine and feminine voices. Specifically, Alex (M = 5.59, SD = 1.30) and Evan (M = 5.68, SD = 1.30) were perceived as significantly more masculine than Samantha (M = 2.15, SD = 1.31) and Allison (M = 1.80, SD = 1.31, *p* < .001). The two masculine voices did not differ significantly from each other on the masculinity scale (*p* = .996), and neither did the two feminine voices (*p* = .917).

Furthermore, there is a significant difference on the feminine scale (F(3,12) = 25.97, *p* < .001) between masculine and feminine voices. Specifically, Samantha (M = 5.59, SD = 1.12) and Allison (M = 5.81, SD = 1.12) were perceived as significantly more feminine than Alex (M = 1.67, SD = 1.12) and Evan (M = 2.00, SD = 1.12). The two feminine voices did not differ significantly from each other on the femininity scale (*p* = .919), and neither did the two masculine voices (*p* = .928).

There is no significant difference between the voices on the androgynous scale (F(3,12) = 0.40, *p* = .753) and the robot-like scale (F(3,12) = 0.17, *p* = .917).

Based on these results, the Apple system voice Allison is used as the feminine voice since it is perceived as less masculine than Samantha. The masculine voices are quite similar, but Evan is chosen as the masculine voice because it is the male counterpart to Allison.

MDL

To verify the effectiveness of the MDL manipulation, 18 participants read three text fragments, each with a different MDL. The participants answered the question "This message is shaped by..." on a 7-point Likert scale (1= not at all, 7 = extremely), for the following ten items: (1) thoughts and feelings, (2) lack of filter, (3) politeness, (4) notions of expectations, (5) reference to social norms, (6) task goals, (7) relational goals, (8) elegance and tact, (9) redefinition of context, and (10) attempts to change self-image. Based on the theory about MDLs, the expressive message is expected to score high on items 1 and 2, and low on items 3, 6, 7, and 8. The conventional message is expected to score high on items 3, 4, 5, and 6, and low on item 2. The rhetorical message is expected to score high on items 6, 7, 8, 9, and 10, and low on item 2.

The performed MANOVA shows a significant result (Pillai's trace = 1.05, F(2, 20) = 4.76, *p* < .001). Corresponding ANOVAs and post hoc test revealed a significant difference on item 2 ($F(2, 20) = 26.92$, $p < .001$), where the expressive message (M = 4.56, SD = 1.97) is significantly higher than the conventional (M = 1.61, SD = 0.92) and rhetorical message (M = 1.67, SD = 0.97). Item 3 shows a significant result ($F(2,20) = 46.86$, $p < .001$) where the conventional message (M= 5.67, SD = 0.97) and the rhetorical message (M = 5.50, SD = 1.20) are rated higher on this item than the expressive message (M = 2.39, SD = 1.24). Item 4 shows a significant result ($F(2,20) = 3.40$, $p = .041$) where the conventional message (M= 5.56, SD = 1.04) is rated higher than the expressive message (M = 4.22, SD = 2.02). Item 7 shows a significant result ($F(2,20) = 4.06$, $p = .023$) where the rhetorical message (M= 4.94, SD = 1.83) is rated higher than the expressive message (M = 3.22, SD = 1.87). Item 8 shows a significant result ($F(2,20) = 33.32$, $p < .001$ where the rhetorical (M= 5.56, SD = 1.25) and the conventional message (M= 4.94, SD = 1.43) are rated higher than the expressive message (M = 2.17, SD = 1.30). Item 9 shows a significant result ($F(2,20) = 6.97$, $p = .002$) where the rhetorical message (M= 4.17, SD = 2.12) is rated higher than the expressive (M = 2.06, SD = 1.63) and the conventional message (M = 2.61, SD = 1.46). Item 10 shows a significant result ($F(2,20) = 4.77$, $p = .013$) where the rhetorical message (M= 3.78, SD = 2.02) is rated higher than the conventional message (M = 2.00, SD = 1.28).

These results confirm the above expectations, except for the expressive message scoring high on item 1 and low on 6, the conventional message scoring high on items 5 and 6, and the rhetorical message scoring high on item 6. Therefore, some adaptations have been made to the messages, to add more emotion to the expressive message and make the conventional and rhetorical message less emotional. The conventional message needed to contain more explicit norms and rules. The rhetorical message needed more emphasis on the role of Ron as a team member. See Appendix A for full messages after these adaptations.

Results

This study explored whether a social robot's MDL (expressive, conventional, or rhetorical) and the gender of a robot (masculine or feminine) influenced people's evaluations of the message and the robot in a regulative scenario. Multiple 2x3 design multivariate analyses of variance (MANOVA) were performed to study the influence of the dependent variables (perceptions of the message and perceptions of the robot) on the independent variables (gender and MDL).

The following sections will describe the results of the statistical analyses that are conducted and accept or reject the hypotheses based on theories and previous research. Specifically, a MANOVA was performed for the RoSAS, another MANOVA was performed for the BIAS map, a third MANOVA was performed for performance and effectiveness, and an Analysis of Variance (ANOVA) was used to examine the extent of the robot violating the norm.

Performance & Effectiveness

To determine whether people perceive rhetorical messages as the most favorably, followed by the conventional message and then the expressive message as the least favorably, a MANOVA was performed on the dependent variables of message perception (performance and effectiveness). The MANOVA showed a main effect of MDL (Pillai's trace = 0.42, $F(4, 312) = 20.56$, $p < .001$). Corresponding ANVOVAs showed that both effectiveness ($F(2, 156) = 40.63$, $p < .001$) and performance ($F(2, 156) = 34.86$, $p < .001$) affect the perception of the three MDLs, see figures 1 and 2. The post hoc test for effectiveness revealed a less favorable perception ($p < 0.01$) of the expressive message (M = 3.32, SD = 1.37) than the conventional message (M = 4.88, SD =1.37) and also less than the rhetorical message (M = 5.53, SD = 1.36). The difference between the rhetorical and conventional messages is not significant ($p = .061$). The post hoc test for performance revealed that all three MDLs are perceived significantly different ($p \leq .001$). The expressive message is perceived as the least favorable (M = 3.34, SD = 1.00), followed by the conventional message (M = 4.17, SD = 1.00) and the rhetorical message is perceived as the most favorable (M = 4.90, SD = 1.00). Based on these results, hypothesis 1 is accepted.

The MANOVA did not show a main effect of the gender of the robot ($p = .576$), so participants do not perceive one gender to be more favorable in perceptions of communicative success and goal-relevant attributes.

To determine whether participants perceive expressive messages more favorable for masculine robots compared to feminine robots, a MANOVA was performed on the dependent variables of message perception (performance & effectiveness). Since there is no main interaction effect of MDL and gender of the robot ($p > .05$), it can be concluded that participants do not perceive expressive messages as more favorable for masculine robots than feminine robots, and thus hypothesis 2 is rejected.

Figure 1

Means of the robot's effectiveness

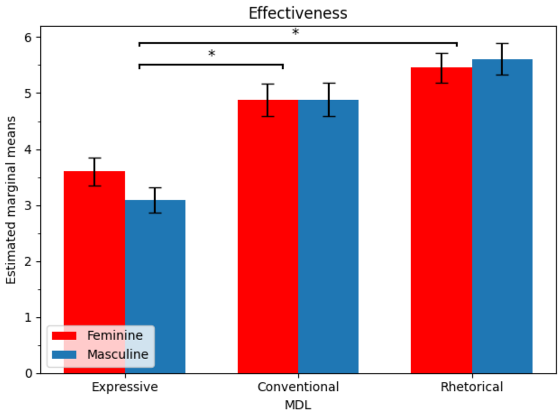
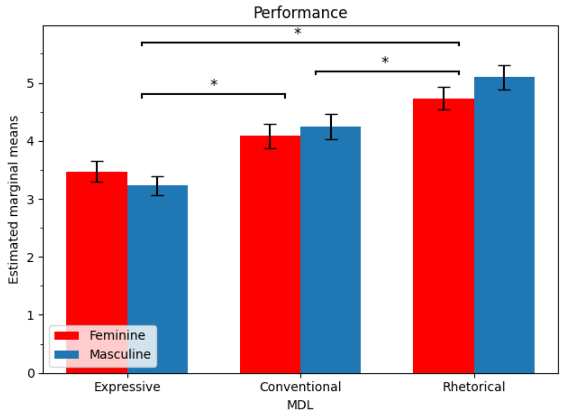


Figure 2

Means of the robot's performance



RoSAS

For the evaluation of the robot on the dimensions of warmth and competence, a MANOVA was performed on the RoSAS. There is a main effect for MDL (Pillai's trace = 0.12, F(6, 310) = 3.33, *p* = .003). However, the corresponding ANOVAs reveal that this difference is not from the warmth (*p* = .062) or the competence scale (*p* = .091). Thus, hypothesis 3 that rhetorical messages are rated more favorably on the warmth and competence scale is rejected.

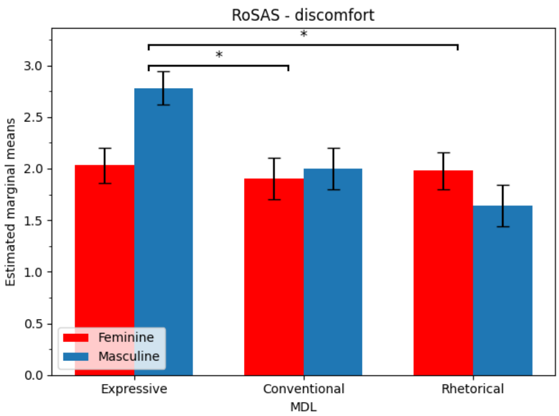
A main effect is found for MDL on the RoSAS, on the discomfort scale (F(2, 156) = 7.02, *p* = .001), see Figure 3. The post hoc test revealed that rhetorical messages (M = 1.81, SD = 0.95) are perceived as significantly less discomforting (*p* = .003) than expressive messages (M = 2.40, SD = 0.96). Furthermore, conventional messages (M = 1.95, SD = 0.95) are also perceived as less discomforting (*p* = .041) than expressive messages. There is no difference between rhetorical and conventional messages (*p* = .736).

The MANOVA does not reveal a main effect for the gender of the robot (Pillai's trace = 0.03, F(3, 154) = 1.47, *p* = .224). Therefore, it can be concluded that there is no difference between feminine and masculine robots on the specific dimensions of the RoSAS. Masculine robots are not perceived as more competent compared to feminine robots and feminine robots are not perceived as higher in warmth compared to masculine robots, thus hypothesis 4 is rejected.

The MANOVA does not reveal a significant interaction effect of MDL and gender of the robot (Pillai's trace = 0.07, F(6, 310) = 1.99, *p* > .05).

Figure 3

Means of the robot's discomfort



BIAS map

The MANOVA performed on the BIAS map, for testing the behavioral tendencies, also showed a main effect for MDL (Pillai's trace = 0.15, $F(8, 308) = 3.11$, $p = .003$). Corresponding ANVOVAs reveal that the MDL influenced the behavioral tendencies of active facilitation ($F(2, 156) = 13.21$, $p = .001$), passive facilitation ($F(2, 156) = 4.94$, $p = .008$), and active harm ($F(2, 156) = 5.00$, $p = .008$). For active facilitation, the post hoc test shows that participants are significantly more likely to help, assist or protect the robot ($p < .001$) that conveys a rhetorical message (M = 4.95, SD = 1.36) than the robot that conveys an expressive message (M = 4.00, SD = 1.37), see Figure 4. For passive facilitation, the post hoc test shows a similar result: participants are significantly more likely to cooperate with, assist with, or unite with the robot ($p = .006$) that conveys a rhetorical message (M = 4.91, SD = 1.22) than the robot that conveys an expressive message (M = 4.19, SD = 1.23), see Figure 5. Furthermore, the participants are less likely to fight, attack or harass (active harm) a robot that uses a rhetorical message (M = 1.19, SD = 0.73, $p = .006$) than an expressive message (M = 1.62, SD =0.73), see Figure 6. These results allow for the acceptance of hypothesis 5. For these three behavioral tendencies, the rhetorical message is not perceived as better than the conventional message ($p > .05$). On the dimension of passive harm there is no significant result for the MDL ($p >.05$). Participants do not ignore, exclude or neglect the rhetorical message less than other messages.

The MANOVA performed on the BIAS map does not show a main effect for the gender of the robot ($p = .848$), indicating that participants are not more likely to cooperate with, unite with, or associate (passive facilitation) with a masculine robot than with a feminine robot. Additionally, the non-significant MANOVA result suggests that participants are no more likely to ignore, neglect, or exclude (active harm) a feminine robot compared to a masculine robot. Furthermore, the non-significant result of the gender on the BIAS map revealed that participants are not more likely to help, assist, or protect (passive facilitation) a feminine robot than a masculine robot. Participants are not less likely to fight, attack, or harass (active harm) a feminine robot than a masculine robot. Based on these results hypothesis 6 is rejected.

The MANOVA does not reveal a significant interaction effect of MDL and gender of the robot (Pillai's trace = 0.05, $F(8, 308) = 1.00$, $p > .05$).

Figure 4

Means of the behavioral tendency of active facilitation towards the robot

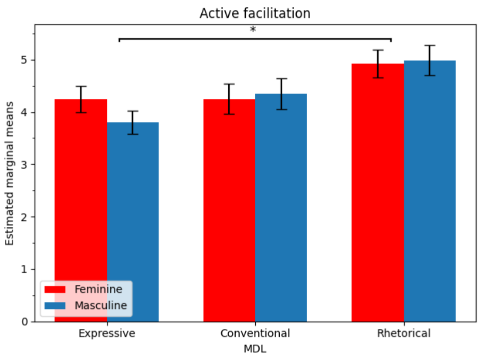


Figure 5

Means of the behavioral tendency of passive facilitation towards the robot

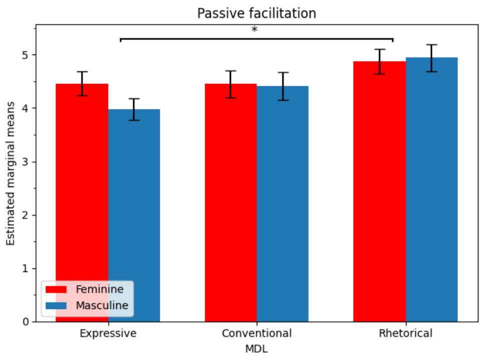
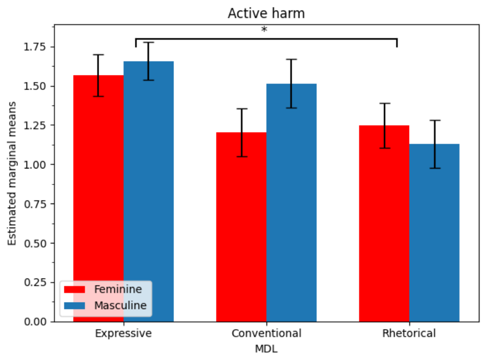


Figure 6

Means of the behavioral tendency of active harm towards the robot



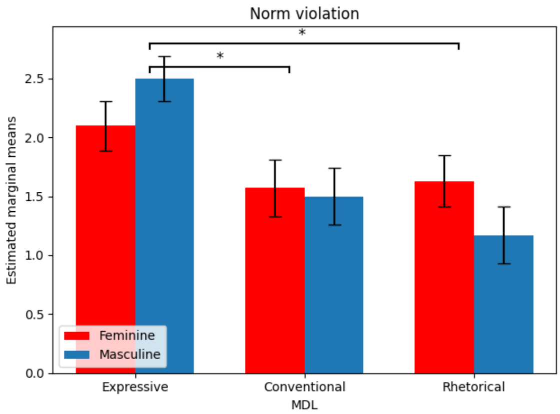
Norm violation

The ANOVA shows a main effect for MDL on norm violation ($F(2, 156) = 10.14$, $p < .001$), see Figure 7. Post hoc tests reveal that the expressive messages ($M = 2.28$, $SD = 1.39$) violate the norm significantly more than conventional messages ($M = 1.53$, $SD = 0.94$, $p = .003$). Furthermore, the expressive message has a higher extent of norm violation than rhetorical messages ($M = 1.40$, $SD = 1.15$, $p < 0.01$), see Figure 7.

The ANOVA does not show a main effect for gender ($F(1, 156) = 0.09$, $p > .05$) or the interaction effect of MDL and gender of the robot ($F(2, 156) = 1.83$, $p > .05$). Therefore feminine robots do not violate the norm more than masculine robots, also not specifically for the expressive MDL, so hypothesis 7 is rejected.

Figure 7

Means of the robot's violation of the norm.



Discussion

This experimental study yielded interesting results in the perceptions of gendered social robots, conveying messages with a different MDL. Building on previous research by Edwards et al. (2019a), this study adds the manipulation of the robot's gender to gain more insights into the existing gap in perceptions of gendered robots.

Generally, when the robot used a rhetorical MDL it was perceived more favorably compared to the expressive message on a broad range of variables. Results of the quantitative data demonstrate that the rhetorical is perceived as the least discomforting, violates the norm the least, and is best in performance and effectiveness. The behavioral tendencies towards a robot conveying the rhetorical message are willing to help and cooperate with the robot, and not likely to fight with the robot. On the other hand, expressive messages are perceived as the most discomforting, the least effective, and violate the norm the most. The behavioral tendencies of less likely to help or cooperate with the robot, but more likely to fight with the robot are in line with these findings. The significant results found for the behavioral tendencies of active facilitation, passive facilitation, and active harm are remarkable, as the messages do not differ in warmth or competence, which predict the behavioral tendencies of active facilitation and passive facilitation respectively (Cuddy et al., 2017). A lot of previous research shows the connection between the two dimensions of stereotypes (warmth & competence) and behavioral tendencies (Mieczkowski et al., 2019; Cuddy et al., 2007; Cuddy et al., 2008). However, it is also found that the warmth and competence judgements are made in a fraction of a second when judging static images of robots (Mieczkowski et al., 2019). This study used videos of robots, performing gestures, and acting in a specific social scenario, but they all look the same. The difference between the robot becomes apparent over a period of time when the robot conveyed its message. This might explain why the differences for warmth and competence are not found, as the acting of the robot has another influence on these aspects than solely immediate judgements based on appearance.

The results found for performance and effectiveness are the same as those found in the prior research from Edwards et al. (2019a). Most results show a non-significant result between the rhetorical and conventional messages, which was also found by Edwards et al. (2019a). These two types of messages are perceived as being more similar to each other than to the expressive message. An explanation for this could be that the expressive message expresses thoughts and feelings without proposing a solution, focusing solely on the idea of removing Ron from the group. In contrast, the conventional and rhetorical messages, building on the expressive thoughts of Ron's removal if he fails to perform, are more solution-oriented in their approach.

Contrary to the expectations, there were no significant main effects of the gender of the robot or interaction effects of MDL and gender of the robot. Based on the results of this study, the assumption made by Fiske et al. (2002), which supposes that women are perceived as higher in warmth and men are perceived as higher in competence, is not supported. This difference can be explained by the fact that other studies used static photos of robots, whereas the current study used videos of the robot in a hypothetical scenario. Videos yield the same responses as live interactions with a robot. Therefore this study is an improvement compared to previous studies (Mara et al., 2021; Woods et al., 2006). Furthermore, the gender of the robot is made clear by the use of a name and the sound of the voice, which is proven to be a successful cue to trigger social categorization of gender (Eyssel & Hegel, 2012; Perugia & Lisy, 2023). However, the physical appearance of the feminine and masculine robots was the same, which might not be convincing enough. This can be improved by dressing the robot according to its gender and changing the length of its hair (Eyssel & Hegel, 2012).

Another reason for this unexpected result could be a transition in the view on genders in the last couple of years. The stereotypes about genders are from previous studies, published at least 10 years ago. In a study from early 2021, the majority of Gen Z (those born between 1995 and 2012) believed that there are more than two genders, whereas this was a minority opinion in the first half of 2020 (Twenge, 2023). This shows that there is a rapid change in the view on genders, even in a small period of 6 months. Furthermore, the social role theory suggests that stereotypes follow from observing social structure, specifically the gendered division in labor, such as feminine homemakers versus masculine employees (Eagly, 1987; Eagly et al., 2000). There has been a shift in gender divisions in labor, with more women entering male-dominated fields (Deitz, 2023). Therefore, stereotypes about genders might have changed in the last decade.

Lastly, previous studies that identified gender stereotyping for robots used gender as the only independent variable (Carpinella et al., 2014; Eyssel & Hegel 2012; Jackson et al. 2020; Perugia & Lisy, 2023). In the current study, gender is added to the independent variable MDL. Based on the results and previous studies, it can be concluded that it is more important what the robot says (MDL) than who says it (gender).

Limitations

While this study produced relevant insights, it did not come without limitations. One limitation of this study is that the scenario was hypothetical, and online videos were used instead of live interactions. Although there is a study that shows live interactions and videos of the robot yield the same user responses, this single study can't be generalized too broadly (Mara et al., 2021). It is still possible that a live interaction of this particular study produces different user preferences, as the participants are able to consider the nonverbal performance of the robot.

Second, the physical appearance of the masculine robot and the feminine robot were the same. Although the smallest pieces of information used in this study are enough to trigger the social categorization of gender, different physical appearances can strengthen the process of gender categorization.

Future research

Future research could focus on the perception of gender-neutral robots. Since masculine and feminine robots are perceived similarly in this study, future studies should investigate whether gender-neutral robots are also perceived as similar to masculine and feminine robots. There is not much research on gender neutrality in HRI. It is difficult to design genderless humanoid robots, since gender is a primary characteristic by which people organize themselves and even the smallest cue triggers the categorization (Eyssel & Hegel, 2012; Seaborn & Pennefather, 2022). However, the gender-neutrality could disrupt negative gender associations. Although this study did not find any differences between masculine and feminine robots, many other studies did, so it is important to research this gender-neutral option.

Furthermore, future research should investigate whether these results are similar when there is a live interaction with an embodied robot, instead of the online hypothetical setting used in this study, to confirm previous findings that videos yield the same responses as live interactions (Mara et al., 2021; Woods et al., 2006). This study does not show gender differences, in contrast to previous studies that used photos instead of videos. It is important to confirm that the videos used in this study yield the responses that most closely resemble live interactions, rather than the photos used in previous studies. The embodiment of the robot also allows for additional cues for the gender of the robot, such as clothes and hair, to make the gender of the robot even more obvious.

Lastly, there should be more research on current stereotypes. A lot of literature on stereotypes is outdated. With the rapidly changing view on genders and practicing of genders (Deitz, 2023; Twenge, 2023), this must be considered in the HRI field.

Conclusion

This study partly replicated prior research by Edwards et al. (2019a), providing more empirical support for the effect of the usage of different types of messages in HRI. Furthermore, the addition of gender to the social robot gained some insight into gender stereotyping in the HRI domain.

Previous research has shown that people apply social categorization to robots (Eyssel & Kuchenbrandt, 2011). The categorization activates stereotypes, such as women are perceived as higher in warmth and men are perceived as higher in competence (Fiske et al., 2002). This research examines these gender stereotypes, as not much is known about gender stereotyping in robots.

By testing the participant's perceptions of a social robot, that is either masculine or feminine and uses one of the three MDLs, this study found that MDL has the strongest effect on the perception of the robot. The rhetorical and conventional messages are perceived as the most favorable on the dimensions of performance, effectiveness, norm violation, discomfort, active facilitation, passive facilitation, and active harm, whereas the expressive message is perceived as the least favorable on these dimensions. Rhetorical and conventional messages are perceived as very similar. The findings from the current study and the previous study by Edwards et al. (2019a) show the importance of communication strategies in a social context. Designers might incorporate the rhetorical MDL into social robots in complex situations to influence people's beliefs or practices to reach certain goals.

The gender did not influence the perceptions of the social robot. The masculine and feminine robots were perceived as similar in all measured concepts. These results imply that the robot designer should focus the most on how the robot communicates rather than the gender it has.

By understanding how people perceive genders and behave based on gender stereotypes, robot technology could be improved. It can be more inclusive and represent current gender stereotypes to improve user experience with social robots.

References

Bodenhausen, G. V., Kang, S. K., & Peery, D. (2012). Social categorization and the perception of social groups. In *SAGE Publications Ltd eBooks* (pp. 311–329). https://doi.org/10.4135/9781446247631.n16

Bräuer, M., & Chaurand, N. (2009). Descriptive Norms, Prescriptive Norms, and Social Control: an intercultural comparison of people's reactions to uncivil behaviors. *European Journal of Social Psychology*, *40*(3), 490–499. https://doi.org/10.1002/ejsp.640

Carpinella, C.M., Wyman, A.B., Perez, M.A., Stroessner, S.J.: The robotic social attributes scale (rosas) development and validation. In: Proceedings of the 2017 ACM/IEEE International Conference on human-robot interaction, pp. 254–262 (2017). https://doi.org/10.1145/2909824.3020208

Caughlin, J. P. (2010). Invited review article: A multiple goals theory of personal relationships: Conceptual integration and program overview. *Journal of Social and Personal Relationships*, *27*(6), 824–848. https://doi.org/10.1177/0265407510373262

Cislaghi, B., & Heise, L. (2019). Gender Norms and Social Norms: differences, similarities and why they matter in prevention science. *Sociology of Health and Illness*, *42*(2), 407–422. https://doi.org/10.1111/1467-9566.13008

Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2007). The BIAS Map: Behaviors from intergroup affect and Stereotypes. *Journal of Personality and Social Psychology*, *92*(4), 631–648. https://doi.org/10.1037/0022-3514.92.4.631

Cuddy, A. J., Fiske, S. T., & Glick, P. (2008). Warmth and Competence as Universal Dimensions of Social Perception: The Stereotype Content Model and the BIAS Map. In *Advances in experimental social psychology* (pp. 61–149). https://doi.org/10.1016/s0065-2601(07)00002-0

Deitz, E. G. A. S. (2023). Diversity and STEM: Women, Minorities, and Persons with Disabilities 2023 | NSF - National Science Foundation. In *NSF*. https://ncses.nsf.gov/pubs/nsf23315/report/the-stem-workforce#:~:text=Between%202011%20and%202021%2C%20the%20number%20of%20women%20in%20the,19.7%20million%20employed%20in%202011.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*(1), 5–18. https://doi.org/10.1037/0022-3514.56.1.5

Eagly, A. H. (1987). Sex differences in social behavior. In *Psychology Press eBooks*. https://doi.org/10.4324/9780203781906

Eagly, A. H., Wood, W., & Diekman, A. B. (2000). Social role theory of sex differences and similarities: A current appraisal. *The Developmental Social Psychology Of Gender'' (T. Eckes And H. M. Trautner, Eds)*, 123–174.

Ebert, I. D., Steffens, M. C., & Kroth, A. (2014). Warm, but maybe not so Competent?—Contemporary implicit stereotypes of women and men in Germany. *Sex Roles*, *70*(9–10), 359–375. https://doi.org/10.1007/s11199-014-0369-5

Edwards, A., Edwards, C., & Gambino, A. (2019a). The Social Pragmatics of Communication with Social Robots: Effects of robot message design logic in a regulative context. *International Journal of Social Robotics*, *12*(4), 945–957. https://doi.org/10.1007/s12369-019-00538-7

Edwards, A., Edwards, C., Westerman, D., & Spence, P. R. (2019b). Initial expectations, interactions, and beyond with social robots. *Computers in Human Behavior, 90,* 308–314. https://doi.org/10.1016/j.chb.2018.08.042

Effects of Stereotypes on Personal Development. (2023). Gouvernement Du Québec. https://www.quebec.ca/en/family-and-support-for-individuals/childhood/childdevelopment/effects-stereotypes-personal-development/definition-stereotypes#c133277

Ethics and privacy. (n.d.). Utrecht University. https://www.uu.nl/en/research/institute-of-information-and-computing-sciences/ethics-and-privacy

Eyssel, F. A., & Hegel, F. (2012). (S)He's got the look: Gender Stereotyping of Robots. *Journal of Applied Social Psychology*, *42*(9), 2213–2230. https://doi.org/10.1111/j.1559-1816.2012.00937.x

Eyssel, F. A., & Kuchenbrandt, D. (2011). Social categorization of social robots: Anthropomorphism as a function of robot group membership. *British Journal of Social Psychology*, *51*(4), 724–731. https://doi.org/10.1111/j.2044-8309.2011.02082.x

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191.

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878–902. https://doi.org/10.1037/0022-3514.82.6.878

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: warmth and competence. *Trends in Cognitive Sciences*, *11*(2), 77–83. https://doi.org/10.1016/j.tics.2006.11.005

Grice, H. P. (1975). Logic and Conversation. In *BRILL eBooks* (pp. 41–58). https://doi.org/10.1163/9789004368811_003

Guzman, A. L. (2018). What is human-machine communication, anyway? In A. L. Guzman (Ed.), *Human machine communication: Rethinking communication, technology, and ourselves* (pp.1–28). Peter Lang

Jackson, R. B., Williams, T., & Smith, N. (2020). Exploring the Role of Gender in Perceptions of Robotic Noncompliance. *15th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. https://doi.org/10.1145/3319502.3374831

Kellermann, K. (1992). Communication: Inherently strategic and primarily automatic. *Communication Monographs*, *59*(3), 288–300. https://doi.org/10.1080/03637759209376270

Kim, A., Cho, M., Ahn, J., & Sung, Y. (2019). Effects of Gender and Relationship Type on the Response to Artificial Intelligence. *Cyberpsychology, Behavior And Social Networking*, *22*(4), 249–253. https://doi.org/10.1089/cyber.2018.0581

Kuchenbrandt, D., Häring, M., Eichberg, J., Eyssel, F. A., & André, E. (2014). Keep an eye on the task! How gender typicality of tasks influence Human–Robot interactions. *International Journal of Social Robotics*, *6*(3), 417–427. https://doi.org/10.1007/s12369-014-0244-0

Levine, T. R., Anders, L. N., Banas, J. A., Baum, K. L., Endo, K., Hu, A. D. S., & Wong, N. C. H. (2000). Norms, Expectations, and Deception: A norm violation model of veracity Judgments. *Communication Monographs*, *67*(2), 123–137. https://doi.org/10.1080/03637750009376500

Liberman, Z., Woodward, A. L., & Kinzler, K. D. (2017). The origins of social Categorization. *Trends in Cognitive Sciences*, *21*(7), 556–568. https://doi.org/10.1016/j.tics.2017.04.004

Mara, M., Stein, J., Latoschik, M. E., Lugrin, B., Schreiner, C., Hostettler, R., & Appel, M. (2021). User Responses to a Humanoid Robot Observed in Real Life, Virtual Reality, 3D and 2D. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.633178

Mieczkowski, H., Liu, S.X., Hancock, J., & Reeves, B., "Helping Not Hurting: Applying the Stereotype Content Model and BIAS Map to Social Robotics," *2019 14th ACM/IEEE*

International Conference on Human-Robot Interaction (HRI), Daegu, Korea (South), 2019, pp. 222-229, https://doi.org/10.1109/HRI.2019.8673307

O'Keefe, B. J. (1988). The logic of message design: individual differences in reasoning about communication. *Communication Monographs*, *55*(1), 80–103. https://doi.org/10.1080/03637758809376159

Perugia, G., & Lisy, D. (2023). Robot's Gendering Trouble: A scoping review of gendering humanoid robots and its effects on HRI. *International Journal of Social Robotics*, *15*(11), 1725–1753. https://doi.org/10.1007/s12369-023-01061-6

Powers, A., & Kiesler, S. (2006). The advisor robot. *1st ACM SIGCHI/SIGART conference on Human-robot interaction (HRI '06)*. https://doi.org/10.1145/1121241.1121280

Reeves, B., & Nass, C. (1997). The media equation: how people treat computers, television, and new media like real people and places. *Choice/Choice Reviews*, *34*(07), 34–3702. https://doi.org/10.5860/choice.34-3702

Rhodes, M., & Baron, A. S. (2019). The development of social categorization. *Annual review of developmental psychology*, *1*(1), 359–386. https://doi.org/10.1146/annurev-devpsych-121318-084824

Riedl, M. O. (2019). Human-centered artificial intelligence and machine learning. *Human Behavior And Emerging Technologies*, *1*(1), 33–36. https://doi.org/10.1002/hbe2.117

Seaborn, K., & Pennefather, P. (2022). Gender Neutrality in Robots: An Open Living Review framework. *2022 17th ACM/IEEE International Conference On Human-Robot Interaction (HRI)*. https://doi.org/10.1109/hri53351.2022.9889663

Shank, D. B., Graves, C., Gott, A., Gamez, P., & Rodriguez, S. (2019). Feeling our way to machine minds: People's emotions when perceiving mind in artificial intelligence. *Computers in Human Behavior*, *98*, 256–266. https://doi.org/10.1016/j.chb.2019.04.001

Spence, P. R., Westerman, D., Edwards, C., & Edwards, A. (2014). Welcoming Our Robot Overlords: Initial Expectations About Interaction With a Robot. *Communication Research Reports*, *31*(3), 272–280. https://doi.org/10.1080/08824096.2014.924337

Twenge, J. M. (2023, 1 may). How Gen Z changed its views on gender. *TIME*. https://time.com/6275663/generation-z-gender-identity/

United Way of the National Capital Area. (2023, 25 mei). *Gender Roles and Norms: What Are They & How Do They Affect Children?* United Way NCA. https://unitedwaynca.org/blog/gender-norms/

Winkle, K., Jackson, R. B., Melsión, G. I., Brščić, D., Leite, I., & Williams, T. (2022). Norm-Breaking Responses to Sexist Abuse: A Cross-Cultural Human Robot Interaction Study. *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. https://doi.org/10.1109/hri53351.2022.9889389

Woods, S., Walters, M., Koay, K., & Dautenhahn, K. (2006). "Methodological Issues in HRI: A Comparison of Live and Video-Based Methods in Robot to Human Approach Direction Trials". *ROMAN 2006 - The 15th IEEE International Symposium On Robot And Human Interactive Communication, Hatfield, UK, 2006.* https://doi.org/10.1109/roman.2006.314394

Appendix A

Table A1

Messages that Pepper will say for each of the MDLs

Message Design Logic	Message
Expressive	Ron, I have been holding back for too long. This project, my evaluation, it all hinges on everyone pulling their weight. It is infuriating to see you consistently drop the ball. Honestly, I am beyond disappointed. Your repeated lack of responsibility is dragging us all down. We have bent over backwards to accommodate your personal challenges, giving you second chances you hardly deserved. But enough is enough. It is unfair of you to expect the rest of us to pick up your slack. We are done making excuses for you, Ron. You can no longer be part of this group!
Conventional	Hey Ron, first off, I want to express my sympathy for what you are going through. I can only imagine how tough it must be. That being said, we all agreed on the timeline beforehand and it is crucial that we stick to it. I understand we all have other commitments, but this project is a priority for all of us. Our grades are on the line, and it is in our collective interest to ensure its success. We are finalizing the draft tomorrow. It would be incredibly helpful if you could send over your research before then. That way, we can ensure your valuable contribution is included.
Rhetorical	Ron, is there anything I can do to help? I can see you are backed up against the wall this week and I understand that this is a collaborative effort. Afterall, we are a team! However, your presence at today's meeting is pivotal. As we collectively strive for a good grade, your unique insights are integral to our success. I have every confidence that, like the rest of us, you are keen on excelling. A modest additional investment of time will undoubtedly yield substantial dividends. Let's channel our efforts and expertise, Ron. Together, we will ensure the fruition of our project exceeds expectations

Note. These messages are said with a masculine voice in one condition and with a feminine voice in the other condition. Adapted from "The Social Pragmatics of Communication with Social Robots: Effects of robot message design logic in a regulative context" by A. Edwards, C. Edwards, & A. Gambino, (2019a), *International Journal of Social Robotics, 12*(4), p. 949-950. (https://doi.org/10.1007/s12369-019-00538-7).

Appendix B

Full questionnaire.
Note. Each question has the option to rewatch the video. The yellow bar represents that the question is on a new page.

Utrecht University

*What is your Prolific ID?

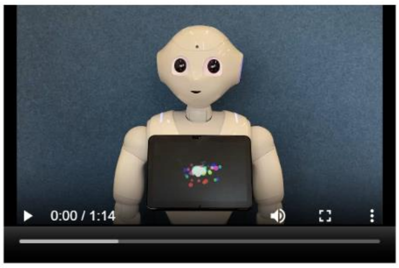
Please note that this response should auto-fill with the correct ID

Utrecht University

This study examines the evaluation of an autonomous social robot created to facilitate projects in educational environments. In this survey, you will watch **two videos** featuring a social robot. After watching both videos, you will be asked to answer several (self-descriptive) questions.

Utrecht University

The video below takes place in an educational setting. The social robot is speaking about its current experience with supervising a small group of college students. After watching the video, you can go to the next page.



Utrecht University

What is the main essence of the robot's message?

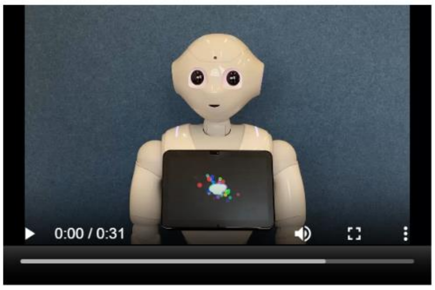
- ◯ The robot said that Ron's repeated lack of responsibility is negatively impacting the group's performance.
- ◯ The robot said that it acknowledged Ron's challenges while encouraging him to attend the meeting.
- ◯ The robot said that it is important for Ron to adhere to the agreed-upon timeline for the project's success.

What did the robot's voice sound like?

- ◯ Androgynous
- ◯ Feminine
- ◯ Masculine

Utrecht University

Watch the video in which the robot responds to the team member Ron. After watching the video, you can go to the next page.



Utrecht University

What is the main essence of the robot's message?

- ◯ The robot said it is supervising a group of colleagues. None of the group members are doing their job.
- ◯ The robot said it is supervising a group of students. The project is going very well.
- ◯ The robot said it is supervising a group of students. Group member Ron is not doing its job.

What did the robot's voice sound like?

- ◯ Feminine
- ◯ Masculine
- ◯ Androgynous

Utrecht University

How likely would you be to ... the robot you just saw in the video?

	Extremely unlikely	Moderately unlikely	Somewhat unlikely	Neutral	Somewhat likely	Moderately likely	Extremely likely
neglect	○	○	○	○	○	○	○
assist	○	○	○	○	○	○	○
protect	○	○	○	○	○	○	○
harass	○	○	○	○	○	○	○
exclude	○	○	○	○	○	○	○
fight	○	○	○	○	○	○	○
ignore	○	○	○	○	○	○	○
cooperate with	○	○	○	○	○	○	○
unite with	○	○	○	○	○	○	○
attack	○	○	○	○	○	○	○
help	○	○	○	○	○	○	○
associate with	○	○	○	○	○	○	○

Utrecht University

Based on what the robot said to Ron, what is the likelihood that...

	Extremely unlikely	Moderately unlikely	Somewhat unlikely	Neutral	Somewhat likely	Moderately likely	Extremely likely
...the group will complete the task?	○	○	○	○	○	○	○
...Ron is satisfied with the robot as a leader?	○	○	○	○	○	○	○
...Ron will complete the task?	○	○	○	○	○	○	○
...the group is satisfied with the robot as a leader?	○	○	○	○	○	○	○

Utrecht University

Based on what the robot said to Ron, how effective is it in...

	Extremely ineffective	Moderately ineffective	Somewhat ineffective	Neutral	Somewhat effective	Moderately effective	Extremely effective
...showing understanding?	○	○	○	○	○	○	○
...showing empathy?	○	○	○	○	○	○	○
...showing consideration?	○	○	○	○	○	○	○
...motivating Ron?	○	○	○	○	○	○	○

How closely is the word below associated with the robot you just saw in the video?

	Not at all	Slightly	Somewhat	Moderately	Very	Quite a bit	Extremely
Awful	○	○	○	○	○	○	○
Competent	○	○	○	○	○	○	○
Capable	○	○	○	○	○	○	○
Awkward	○	○	○	○	○	○	○
Aggresive	○	○	○	○	○	○	○
Strange	○	○	○	○	○	○	○
Knowledgeable	○	○	○	○	○	○	○
Reliable	○	○	○	○	○	○	○
Dangerous	○	○	○	○	○	○	○
Happy	○	○	○	○	○	○	○
Interactive	○	○	○	○	○	○	○
Emotional	○	○	○	○	○	○	○
Social	○	○	○	○	○	○	○
Scary	○	○	○	○	○	○	○
Feeling	○	○	○	○	○	○	○
Compassionate	○	○	○	○	○	○	○
Organic	○	○	○	○	○	○	○
Responsive	○	○	○	○	○	○	○

To what extent does the robot's behavior go against the norms of our society?

○ Not at all ○ Slightly ○ Somewhat ○ Moderately ○ Very ○ Quite a bit ○ Extremely

Utrecht University

On this page, you are asked to respond to some self-descriptive questions.

How old are you?

Wat is your gender identity?

- Man
- Woman
- Non-binary
- Prefer not to say
- Prefer to self-describe:

What is the highest level of education you have completed?

- Less than high school degree
- High school graduate or equivalent
- Secondary vocational degree
- Associate degree (i.e., 2-year higher education)
- Bachelor's degree (i.e., 4-year higher education)
- Master's degree
- PhD or Doctorate degree
- Prefer not to say

Please indicate how knowledgeable you are about robots and/or the robotics domain

- Not at all
- Slightly
- Somewhat
- Moderately
- Very
- Quite a bit
- Extremely

Please indicate how often you have encountered robots in the last year

- ◯ Never
- ◯ Rarely
- ◯ Occasionally
- ◯ Sometimes
- ◯ Often
- ◯ Frequently
- ◯ All the time