4th International Symposium on New Frontiers in Human-Robot Interaction
21–22 April 2015, Canterbury, UK
M. Salem, A. Weiss, P. Baxter, K. Dautenhahn (Eds.)

Gender, more so than Age, Modulates Positive Perceptions of Language-Based Human-Robot Interactions

Megan Strait and **Priscilla Briggs** and **Matthias Scheutz**[1]

Abstract. Prior work has shown that a robot which uses politeness modifiers in its speech is perceived more favorably by human interactants, as compared to a robot using more direct instructions. However, the findings to-date have been based soley on data aquired from the standard university pool, which may introduce biases into the results. Moreover, the work does not take into account the potential modulatory effects of a person's age and gender, despite the influence these factors exert on perceptions of both natural language interactions and social robots. Via a set of two experimental studies, the present work thus explores how prior findings translate, given a more diverse subject population recruited via Amazon's Mechanical Turk. The results indicate that previous implications regarding a robot's politeness hold even with the broader sampling. Further, they reveal several gender-based effects that warrant further attention.

Figure 1: Scenario: the humanoid MDS robot (Xitome Designs; left) instructs a confederate participant (right) on a brief drawing task.

1 INTRODUCTION

Natural language interactions with virtual and robotic agents are becoming increasingly pervasive, from virtual personal assistants (such as Apple's Siri agent), to socially assistive robots (e.g., elder care robots such as [4]). As the functionality of these artificial agents grows, so does the need to communicate with humans effectively to best serve the human interlocutor [12]. Surprisingly, however, there are very few attempts to date to carefully evaluate the different ways in which artificial agents could talk with humans in the context of a given task based on the agent's physical embodiment. For example, it is unclear whether an artifical agent, depending on its embodiment, should use imperatives when instructing humans (e.g., "turn right at the next intersection") or whether a more polite way of expressing an instruction is required (e.g., "we need to turn right at the next intersection"). Intuitively, a non-embodied agent like a navigation system might get away with syntactically simple, effecient imperatives, while a humanlike embodied robotic agent might have to employ more conventional forms of politeness.

Past work evaluating politeness in natural language interactions with robotic agents supports this intuition. Torrey and colleagues, for example, showed that the use of hedges (e.g., "*I guess*", "*probably*", and "*sort of*") and discourse markers – two "negative" politeness techniques – improves how people perceive a robot instructing a person via natural language. Specifically, they found that polite robots were viewed more positively than robots using more direct speech [22]. Even though negative politeness may be less noticeable than the *please*s of positive politeness, hedging indicates to the listener that the speaker is trying to mitigate the force of the request [7, 14].

[1] Human-Robot Interaction Laboratory, Tufts University, Medford, MA USA

Recent extensions of the above findings show that other negative politeness techniques (e.g., phrasing requests indirectly [9]), as well as positive (e.g., inclusive pronouns), suffice to improve perceptions of human-robot interactions (e.g., [6, 19, 21]). However, this research investigating human perceptions of robot politeness in human-robot interactions ([21, 22]) is predominately based on data drawn from the standard (and relatively homogeneous) university population.

Thus, whether and how these findings transfer to scenarios involving a population that is more diverse (e.g., economically, educationally), remains unknown. In particular, there are several factors (socio-linguistic, cultural, and demographic) in addition to politeness that have been found to modulate perceptions of natural language interactions (e.g., [3, 13, 16, 17, 20]). For instance, contrary to popular stereotypes, Japan is not as robot-positive as the US [2, 8].

Of particular relevance, is the growing amount of evidence that men (relative to women) hold significantly more positive towards robotic entities [5]. While both Torrey et al. ([22]) and Strait et al. ([21]) attempted to control for unintended effects due to gender, their participant samples were nevertheless imbalanced and thereby constrained in their ability to represent the general population. Hence, it is important to revisit these findings with explicit consideration of socio-demographic factors to understand what are their specific influences and how the findings extend beyond the university.

The goal of the present work was thus two-fold: (1) to investigate whether an extension of [21] with more diverse subject demographics would replicate the previously-observed effects of robot politeness (based on interaction observation), and further, (2) how the subject-based factors of age and gender specifically interact with those of the robot (e.g., the robot's use of polite communicatory cues).

To address these questions, we conducted a set of two online experiments via Amazon's Mechanical Turk with the aim of achieving greater diversity in people's age, and educational/geographical backgrounds, as well as more balanced gender demographics. In both, we presented videos depicting a robot instructing a person on a simple drawing task. We solicited people's reactions to these videos to determine the influence of a robot's politeness relative to any modulatory effects of a person's age and gender (**Experiment I**). Owing to a limitation of the first study, we conducted a follow-up to Experiment I to determine whether the findings hold given more naturalistic interaction settings (**Experiment II**).

2 EXPERIMENT I

Based on the previous work outlined in the introduction ([21, 22]), we hypothesized that by using politeness modifiers in its speech, a robot would be perceived more favorably (as evidenced by higher ratings of *likeability* and reduced ratings of *aggression*) than a robot that uses more direct instructions. In addition, we generally explored the modulatory effects of a person's socio-demographic factors – in particular, age and gender – and how they interact with characteristics of a robot to influence perceptions of human-robot interactions.

To test our hypotheses and the age- and/or gender-based modulations thereof, we conducted a fully between-subjects investigation of the effects of a robot's communication strategy on observations of brief human-robot interactions – as influenced by a person's age and gender. In order to obtain a more diverse population than previously, we conducted our investigation online via Amazon's Mechanical Turk. Using a modification of the materials and methods developed in [21], we tasked participants with viewing a short video depicting a robot as it advised a person on creating a simple drawing. Following the video viewing, participants were prompted for their perceptions of the interaction, as rated on several dimensions regarding the likeability and aggression of the robot.

2.1 Materials & Methods

2.1.1 Participants & Procedure

839 participants were recruited via Amazon Mechanical Turk.[2] Prior to participating, subjects were informed the purpose of the study was to investigate factors that influence perceptions of human-robot interactions. Upon informed consent and subsequent completion of a demographic survey, the subject was shown one of 32 videos depicting a robot instructing a human confederate on a simple task. Following the viewing, the subject completed a 12-item questionnaire regarding his/her perceptions of the robot's appearance and behavior. Lastly, to assess attentiveness, participants completed a three-item check regarding salient details of the video clip.

Of these 839 participants, data from 329 were discarded due to several exclusion criteria: a restriction to limit participation to native english speakers (51 participants), and failure to complete the requested tasks (70) or failure on a three-item attention check (with a success threshold of 100%) to ensure participants viewed the presented video (208). Thus, our final sample included data from 510 participants (62% male) from 47 of 50 US states. The average age of this sample was 31.21 (SD=9.71), ranging from 18 to 76 years old. The most common level of education obtained was a bachelor's degree (45%), with an additional 36% of participants having some amount of college-level education. A small percent of participants reported having completed only high school (12%) and a smaller proportion reported obtaining more advanced degrees (7%). Participants also reported relatively high interest in robots (M=5.15, SD=1.32) – though low familiarity with robots (M=3.75, SD=1.49) – based on a 7-point Likert scale with 1=low and 7=$high$.

[2] In anticipation of some loss in data due to exclusion criteria, we chose this sample size to achieve ≥ 15 useable observations in hypothesis testing.

	Comforting	*Considerate*	*Controlling*
Aggressive	−.15	−.11	**.68**
Annoying	**-.62**	−.27	.21
Comforting	**.73**	.30	−.13
Considerate	.21	**.63**	−.15
Controlling	−.11	−.16	**.52**
Eerie	**-.73**		.16
Likable	**.60**	**.59**	
Warm	.22	**.77**	−.24
Eigenvalues	3.63	1.16	.99
Variance Explained	.24	.44	.56

Table 1: Factor loadings for the three-factor EFA solution.

2.1.2 Independent Variables

We employed a $2 \times 3 \times 2$ factorial design in which we systematically manipulated a robot's *politeness* in an advice-giving scenario, using the same conditions as those developed by Strait and colleagues ([21]). We also included participant *age* (three levels) and *gender* to investigate how they affect perceptions of the human-robot interaction. In total, we had the following three independent variables (IVs):

- **Politeness** of the robot's instructions (direct vs. polite). The *polite* condition entailed the robot giving instructions that contained one or more of both positive and negative politeness strategies, such as praise (e.g., "great job") and hedges (e.g., "a *kind of* large circle"). The *direct* speech condition employed the exact same instructions, but with the politeness modifiers removed.
- **Participant age** (three levels). We established three age categories based on a $1/3$ split of all the self-reported ages, resulting in a corresponding to the age of the standard university sample (M_1=22.81 years, SD=1.87), as well as two older adult categories (M_2=28.68, SD=1.99; M_3=42.16, SD=8.86).
- **Participant gender** (female vs. male).

2.1.3 Covariates

In addition to the above, we planned to carefully control for potential effects due to a person's motivations for completing the tasks (i.e., due to his/her purported *interest* in robots), as well as any effects due to characteristics of the stimulus set. To do so, we covaried three factors pertaining to the robot's physical embodiment:

- **Appearance** of the robot (two levels): the humanoid MDS (Xitome Designs) versus the less humanlike PR2 (Willow Garage).
- **Production modality** (synthetic vs. human speech), and
- **Gender** (female vs. male) of the robot's voice.

Thus, a total of four covariates – participants' **interest** in robots, the robot's **appearance** and the **gender** and **production modality** of the robot's voice, – were used in the analyses reported below.

2.1.4 Stimuli

A set of 32 videos (two conditions – polite versus direct speech – with 16 instances per condition) were constructed based on systematic manipulation of the robot-based IVs and covariates. Each video depicted a variant of a robot instructing a male human actor on a pen-and-paper drawing of a koala (cf. [21]). To avoid potential effects of affect, behavior, and/or movement (due to differences between the two robots' abilities), the robots were kept stationary. To avoid unintended effects due to a particular appearance, gender, voice, or the way in which the voice was produced, 16 video instances co-varying the robot's humanoid appearance (MDS versus the PR2), voice production modality (synthetic- versus human-produced speech) and voice gender (four voices – two female, two male) were created per condition. Four adult human actors comprised the set of human voices, with instructions to perform with flat affect. Synthetic voice production was performed using the native Mac OS X text-to-speech (TTS) software with four voices: "Alex","Ava", "Tom", and "Vicki". Following a between-subjects design, participants viewed only one video (selected randomly from the set of 32).

2.1.5 Dependent Variables

Of the set of 12 questionnaire items, three items – *task difficulty*, *interaction difficulty*, and *interest in interacting* – were considered as unique variables. On the remaining 9 items drawn from prior work (cf. [21, 22]), exploratory factor analysis produced a three-factor solution which showed a better fit ($\chi^2(7) = 13.36$, $p = .0638$) than a model where the variables correlate freely.

The criterion for retention of a questionnaire item was a factor loading of $\geq .50$ (see Table 1). We thus interpreted the three latent variables as the following: how **comforting** (four items – comforting, likable, -annoying, and -eerie; Cronbach's α=.83), **considerate** (three items – considerate, likable, and warm; α=.79), and **controlling** (two items – aggressive and controlling; α=.55) the robot was perceived. Items that were negatively correlated are indicated by $-$, and were automatically reversed in the computation of the latent constructs. Further, all dependent measures were normalized (to a scale between 0 and 1) prior to analysis.

2.2 Results

To assess the effects of the three IVs, between-subjects ANCOVAs were conducted on each of the dependent variables (taking into account the four covariates), with homogeneity of variance confirmed using Levene's test. All significant effects are reported below (with significance denoting $\alpha \leq .05$), and all post-hoc tests reflect a Bonferroni-Holm correction for multiple comparisons.

2.2.1 Comforting, Considerate, & Controlling

As expected, the *politeness* manipulation showed marginal ($p<.10$) to significant main effects on all three latent factors – *comforting*, *considerate*, and *controlling* (see Table 2, top). Similarly, participants' *gender* did as well (see Table 2, bottom); however, there were no significant main or interaction effects due to the participants' *age*.

Overall, both politeness and gender tended to increase ratings of the robot as *comforting* and *considerate*, and conversely, decrease those for *controlling*. However, these main effects were eclipsed by a *politeness* $\times$ *gender* interaction on both of the two positive factors: *comforting* ($F(1,498)$=4.57, p=.03, η^2=.01) and *considerate* ($F(1,498)$=6.97, p<.01, η^2=.01).

	DIRECT ($n = 254$)	**POLITE** ($n = 256$)	$F(1,498)$	p	η^2
Comforting	.13 (.37)	.19 (.38)	3.26	$= .07$	.01
Considerate	.46 (.16)	.54 (.17)	31.82	$< .01$	**.06**
Controlling	.25 (.17)	.20 (.16)	10.29	$< .01$	.02

	FEMALE ($n = 193$)	**MALE** ($n = 317$)	$F(1,498)$	p	η^2
Comforting	.21 (.40)	.11 (.36)	9.27	$< .01$	.02
Considerate	.53 (.16)	.48 (.16)	13.42	$< .01$	.03
Controlling	.20 (.16)	.26 (.17)	14.44	$< .01$	.03
Difficulty (t)	.17 (.16)	.21 (.18)	8.20	$< .01$	.02
Difficulty (i)	.24 (.23)	.28 (.21)	5.18	$= .02$	.01
Interest	.48 (.23)	.43 (.21)	4.74	$= .01$	.01

Table 2: Main effects of *politeness* (top) and *gender* (bottom), and relevant descriptive and inferential statistics.

In particular, the interactions showed that – while polite speech tended to improve participants' ratings – it did so primarily for women (see Figure 2, left and center). That is, a robot's use of polite speech significantly improved ratings of *comfort* when viewed by female observers (M=.29, SD=.39, n=94) relative to those by female observers of direct speech (M=.14, SD=.40, n=99; p=.04) and male observers of both direct (M=.11, SD=.34, n=155; p=.01) and polite speech (M=.11, SD=.38, n=162; p=.01). Similarly, though the polite robot significantly improved observers' ratings of *considerateness* for both female (M_{polite}=.59, SD=.16; M_{direct}=.47, SD=.17; p<.01) and male observers (M_{polite}=.50, SD=.17; M_{direct}=.45, SD=.15; p=.02), women's ratings were most improved relative to men's (p<.01).

With regard to perceptions of the robot as *controlling*, politeness was still broadly effective at decreasing ratings – regardless of the observer's gender, with polite robots receiving lower ratings relative to those more direct in their instructions (see Table 2, top). But, just being female helped as well: with women rating the robot as substantially less controlling than did men (see Table 2, bottom).

2.2.2 Difficulty & Interest

Gender further exerted significant main effects on the dependent variables regarding the perceived *difficulty* of both the task and interaction, as well as the observers' own *interest* in interacting with the depicted robot (see Table 2, bottom). In particular, female participants tended to rate both the task and interaction as less difficult than did males (see Table 2–bottom, *Difficulty*). Furthermore, they tended to show more interest in interacting with the robot agent than their male counterparts (see Table 2–bottom, *Interest*). There were no significant effects (main or interaction) due to *politeness* or *age*.

2.3 Discussion

Do people perceive a robot, which employs politeness modifiers in its speech, more favorably than one that uses more direct speech? Based on previous research by [21, 22], we expected that participants would rate a polite robot more favorably than one that is more direct in its instructions, as evidenced by higher ratings of positive constructs (e.g., *likability*) and lower ratings of negative constructs (e.g., *aggression*). Consistent with that work, the politeness manipulation here showed lower ratings of the robot as *controlling* and higher ratings of the robot as being *considerate* and *comforting*. In particular, our results

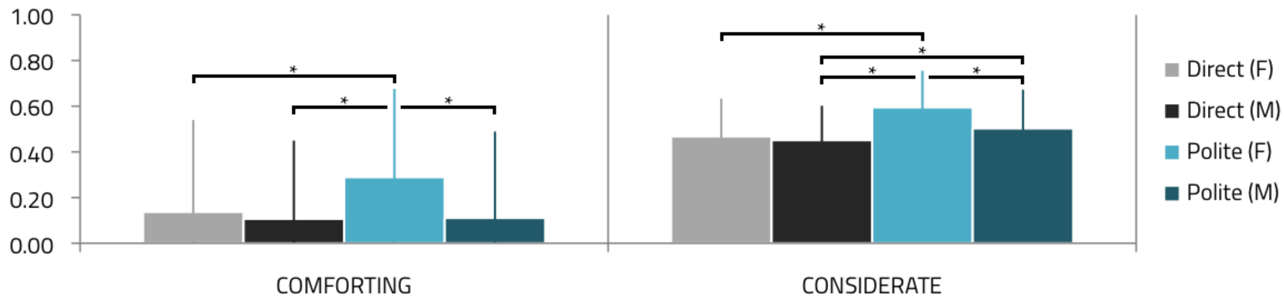

Figure 2: Interaction between robot *politeness* and participant *gender* on the three latent factors – the degree to which the robot was perceived as *comforting*, *considerate*, and *controlling*. Gray bars indicate the use of *direct* speech, versus blue, which indicates *polite* speech. Lighter bars indicate female participants (versus male participants, darker bars). All significant contrasts are shown (indicated by asterisks).

replicate and confirm those of prior work, even with a substantially more diverse subject population.

Does a person's age and gender further modulate perceptions of human-robot interactions? Based on previous suggestions that men and women view and respond to robots in significantly different ways [5], we evaluated the primary and modulatory effects of participants' age and gender. Participants' *gender* exerted a main effect on *all* dependent measures: how *comforting*, *considerate*, and *controlling* the robot was perceived as being, as well as how *difficult* both the task and interaction seemed and participants' *interest* in interacting with the depicted robot. In particular, female participants (relative to their male counterparts) showed more positive responding towards the robots and their interactions with the human confederate, as reflected by increased ratings of interest, comfort, and the robot's considerateness, as well as decreased ratings of the task/interaction difficulty and the robot's aggression. Further, interactions with the politeness manipulation showed that a robot's use of polite speech was effective at increasing women's positive attributions (the robot as being *comforting* and *considerate*), but not men's. Participant *age*, however, showed no main or interaction effects on *any* of the measures.

2.3.1 Implications

Prior work has suggested that a robot's use of politeness modifiers in its speech improves perceptions of human-robot interactions in advice-giving situations [21, 22]. Our results further replicate these findings (with respect to observation of human-robot interactions), and moreover, show the influence of politeness holds given a more general and representative population sample. In particular, our participants came from a wide variety of educational backgrounds (ranging from high school to advanced degrees) and geographical locations within the US (47 states).

In addition, we explicitely considered the effects of a person's age (ranging from the standard university age level to older adult) and their gender, to determine their influence and nature relative to the robot's politeness. This consideration of such socio-demographic factors revealed a methodological consideration for HRI studies – namely, that a person's *gender* should be taken into account when assessing perceptions of language-based human-robot interactions, as it is a modulating influence in addition to a robot's *politeness*.

This was expected, as previous research (e.g., [15, 18, 20]) has found that men exhibit more positivity towards robots than women. But, contrary to prior observations, our results indicate that women respond, in general, more positively towards the depicted robots. This may be due to the difference in the presentation the interactions as, here, video-recordings of human-robot interactions were evaluated by post-hoc observation, whereas, previous work has used scenarios involving the participatory and co-located interaction between the participant and robot of interest [16, 17, 20]. Alternatively (or in addition), it may be due to the difference in interaction: here, the robot interactants were depicted as instructing a human confederate; whereas, the human interactants in prior work were tasked with instructing or working with (rather than subservient to) the robot agent. Despite the conflicting differences in the nature of their effects, our findings add to the growing body of evidence implicating gender as an important methodological consideration in evaluating perceptions of human-robot interactions.

2.3.2 Limitations & Future Directions

Our approach to the investigation of perceptions of polite robots contributes a simple online task to assess the modulatory influences (or lack thereof) of a person's age and gender. In particular, the collection of data with broad socio-demographics augments in-laboratory studies that are limited to small, and relatively homogeneous, participant populations. This contribution here is significant because it replicates the previously reported influences of politeness, and further, sheds light on how such findings might transfer to the general population. That said, our approach also has several limitations (which underscore avenues for further research), three of which we discuss below.

Relevance. First, we note that the effect sizes for the given manipulations are relatively small. The magnitude of the effect of politeness on perceptions of the robot's considerate approaches a medium qualification (η=.10), but nevertheless, the implications of both robot politeness and participant gender are of limited weight. This may also suggest it is worth looking at the specific effects due to other factors such as a person's educational or geographic background (two socio-demographic items for which we did not control).

Mode of Evaluation. Another limitation of relevant consideration is how peoples' evaluations of the interactions were obtained. Here, the interactions were evaluated post-hoc by a third-party observer, who (by definition) was remotely located from the actual robot/interaction. This is particularly important to note, as it has been found that perceptions of human-robot interactions are further modulated by the interaction distance (remote versus co-located) and nature (observatory versus participatory) [21]. Thus, while the video-based interactions and online evaluations allowed us to sample from a broader demographic than that which is available locally, whether and how our gender-based findings apply to actual, co-located human-robot interactions warrants further investigation.

Stimuli. Lastly, there are a number of important limitations to the stimuli used and their presentation. Here the stimuli depicted brief (2 minute) interactions between an inanimate robot and a human confederate, which is an unrealistic interaction scenario in comparison to the intended usage of social robots.

In particular, prior work has shown that movement (however subtle) can impact the efficacy of interactions. For example, Andrist and colleagues have found that averting a robot's gaze (even for robot's without articulated eyes) can improve perceptions of the robot and their interactions [1]. Thus, with regard to the present study – though we limited movement to avoid unintended and/or differential influences (e.g., due to the robots' different capacities for actuation), the absence of movement itself might be affecting the current findings in unknown ways. For instance, the absence of attention-indicating gaze (e.g., looking at the participant when he/she is not performing a drawing instruction) might reduce positive attributions (e.g., considerateness) and/or increase negative attributions. This idea is supported by participants' open responses, which generally showed negative attitudes regarding the robot's lack of movement. Thus, there is the distict possibility that the lack of movement influenced perceptions in some way that may attenuate (or worse, decimate) other influences (e.g., due to politeness). With such considerations in mind, we moved to conduct a follow-up experiment to test the nature and magnitude of effects due to politeness and gender, when the robot was animated in a more naturalistic fashion.

3 EXPERIMENT II

Based on the considerations outlined in the previous section, we composed an exploratory follow-up investigation to Experiment I. We again conducted a between-subjects investigation of the effects of a robot's politeness (as influenced by a person's gender) on perceptions of human-robot interactions – but, with more naturalistic interactions. Specifically, we constructed a second set of video stimuli in which the robot was *animated* with attention-sharing and (human-like) idling movements, based on the naturalistic movements exhibited by a human instructing in such a context.

3.1 Materials & Methods

3.1.1 Participants & Procedure

437 additional participants were recruited via Amazon Mechanical Turk.[3] As in Experiment I, participants were told the purpose of the study was to investigate factors that influence perceptions of human-robot interactions. Upon informed consent and completion of a demographic questionnaire, the subject was shown one of 16 videos (similarly depicting a robot instructing a human confederate on a simple task). Following the viewing, the subject completed the 12-item questionnaire regarding his/her perceptions of the robot's appearance and behavior and the three-item check to assess whether the participant attended to the video.

Of these 437 participants, data from 176 participants were discarded due to: failure to complete the requested tasks (54) or failure on the attention check (122). Thus, our final sample included data from 261 participants (60% male) from 48 of the 50 US states. The average age of this sample was 32.45 (SD=10.45), ranging from 18 to 68 years old. The most common level of education obtained was similarly a bachelor's degree (44%), with an additional 37% of participants having some amount of college-level education. As in Experiment I, a small percent of participants reported having completed only high school (13%) and a smaller proportion reported obtaining more advanced degrees (6%). Participants again reported low familiarity (M=3.79, SD=1.49) with, but relatively high interest (M=5.33, SD=1.39) in robots.

3.1.2 Independent Variables

We again employed a fully factorial design, with the same independent variables as previously:

- The robot's **politeness** (direct vs. polite).
- **Participant age** (three levels): the standard university sample (M_1=22.85 years, SD=2.18), as well as two older adult categories (M_2=29.70, SD=2.01; M_3=43.98, SD=8.65).
- The participant's **gender** (female vs. male).

3.1.3 Covariates

We again planned to control for effects due to a person's *interest* in robots, as well as any due to characteristics of the stimulus set. As there was little variance explained by *production modality*, we excluded it from consideration to help reduce the overall number of videos to remake, thus reducing the number of observations needed to achieve similar sample sizes as Experiment I. As a result, we considered a total of three covariates in our analyses here: two factors pertaining to the robot's physical embodiment (the robot's *appearance* – MDS vs. PR2 – and *gender* of the robot's voice) and one factors pertaining to the participant (their *interest* in robots).

3.1.4 Stimuli

To increase the degree of observable presence/embodiment of the depicted robots, we recreated the videos from Experiment I[4] to animate the robots with select movements during the interaction. The movement modifications were intended to create a sense of "shared attention" and "idle" behaviors, based on the behaviors observed of a human instructor during pretesting of the drawing task with two people. In particular, the attentive behaviors were implemented such that the robot (MDS or PR2) moved its eyes (MDS) or head (PR2) up/down to focus on the human actor when giving instructions or on the actor's drawing (when the actor was drawing). Each robot also performed a set of idle behaviors (initiated based on random timers) throughout the interaction, based on their relative capacities for movement:

- *Blinking* (MDS only) – the MDS robot has two actuated eyelids that were closed and reopened ($500ms$) mimic human blinking.
- *Swaying* (MDS only) – the MDS has three degrees of freedom (DOF) on its center axis, allowing mimicry of slight head tilts (left/right and up/down positioning determined randomly at initiation of each tilt).
- *Breathing* (PR2 only) – the PR2, having fewer DOF with respect to its head movement, was limited to regular up/down undulation of its frontal laser. The rate of the laser movement approximated the average person's resting state heart rate ($70bpm$).

[3] In anticipation of data loss due to our exclusion criteria, we chose this sample size to again achieve ≥ 15 useable observations in hypothesis testing.

[4] As *production modality* was dropped from consideration, we recreated only a subset of the E1 videos – the 16 depicting a robot with a synthetic voice.

	DIRECT ($n = 130$)	POLITE ($n = 131$)	$F(1, 249)$	p	η^2
Comforting	.59 (.20)	.66 (.20)	6.72	= .01	.03
Considerate	.59 (.18)	.70 (.17)	26.27	< .01	**.11**
Controlling	.24 (.19)	.19 (.15)	6.31	= .01	.03

	FEMALE ($n = 104$)	MALE ($n = 157$)	$F(1, 249)$	p	η^2
Comforting	.68 (.20)	.58 (.20)	15.03	< .01	**.06**
Considerate	.68 (.18)	.62 (.18)	8.67	< .01	.03
Controlling	.20 (.16)	.24 (.17)	4.22	= .04	.02
Difficulty (t)	.24 (.23)	.29 (.23)	3.55	= .06	.01
Difficulty (i)	.26 (.28)	.35 (.27)	6.40	= .01	.03
Interest	.68 (.28)	.61 (.27)	3.45	= .06	.01

Table 3: Main effects of *politeness* (top) and *gender* (bottom), and relevant descriptive statistics, in Experiment II.

3.1.5 Dependent Variables

We used the same dependent measures as previously: task and interaction *difficulty* and *interest in interacting*, as well as how **comforting**, **considerate**, and **controlling** the robot was perceived as being.

3.2 Results

To assess the effects of robot *politeness* and participant *age/gender* – in the context of more naturalistic interactions – between-subjects ANCOVAs were conducted on each of the dependent variables (taking into account the four covariates), with homogeneity of variance confirmed using Levene's test. All significant effects are reported below (with significance denoting $\alpha \leq .05$), and all post-hoc tests reflect a Bonferroni-Holm correction for multiple comparisons.

3.2.1 Robot Politeness

As previously found, *politeness* exerted a significant effect on all three of *comforting*, *considerate*, and *controlling* DVs. Specifically, as expected based on Experiment I and previous literature, the robot's use of polite speech increased participants' comfort and their perceptions of the robot's considerateness. It also reduced perceptions of the robot as controlling (see Table 3, top).

3.2.2 Participant Age & Gender

Similarly, as Experiment I showed, *gender* improved perceptions along all dependent measures (see Table 3, bottom). Specifically, female participants continued here to (1) rate the robot as more *considerate* and less *controlling*, (2) indicate greater *comfort* and *interest in interacting* with the depicted robot, and (3) rate both the interaction and task as less difficult, than did their male counterparts.

Unlike the previous experiment, however, here participant *age* also showed a significant influence on *comfort* with the robot ($F(2, 249) = 3.19$, $p = .04$, $\eta^2 = .03$) and perception of it as *controlling* ($F(2, 249) = 4.07$, $p = .01$, $\eta^2 = .03$). Specifically, participants of the standard university age (young adults) indicated significantly less comfort with the robot ($M = .59$, $SD = .22$, $n = 87$) than the oldest participants ($M = .67$, $SD = .19$, $n = 92$; $p < .01$). Conversely, the younger participants also rated the robot as significantly more controlling ($M = .26$, $SD = .19$, $n = 87$; $p = .01$) than did either of the two older age groups – middle adults ($M = .19$, $SD = .16$, $n = 82$) and older adults ($M = .20$, $SD = .16$, $n = 92$).

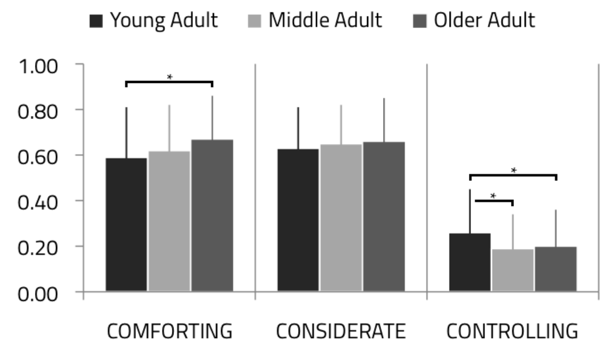

Figure 3: Main effects of participant *age*. Asterisks indicate significant contrasts.

3.2.3 Interactions

Furthermore at odds with Experiment I (where the $gender \times politeness$ eclipsed many of the main effects of politeness), there were no significant or even marginally significant interaction effects here. Specifically, in the context of the more naturalistic interactions, the use of polite speech seemed to be effective for both female *and* male participants. This suggests that, while female participants appear to be particularly sensitive to verbal communication (as evidenced by their ratings across both the more naturalistic Experiment II *and* Experiment I), male participants may be more sensitive to *consistency* in verbal and nonverbal communicatory cues.

3.3 Discussion

3.3.1 Summary of Present Findings & Implications

In this follow-up investigation, we explored whether our previous findings in Experiment I – that a robot's use of polite speech improves perceptions (and, that women respond more positively towards such robots) – hold given more naturalistic interaction scenarios (i.e., human-robot interactions in which the robot is animated).

Here we observed that the results, for the most part, reflect those of the previous experiment (despite E1's lack of movement in the shown video interactions). Specifically, the politeness manipulation again resulted in lower ratings of the robot as *controlling* and higher ratings of the robot as being *considerate* and *comforting* (see Figure 4, left). This lends further support of politeness as an effective tool for facilitating more positive responding towards robots (at least for natural language interactions in advice-giving scenarios).

Similarly, participants' *gender* again exerted a main effect on *all* dependent measures: how *comforting*, *considerate*, and *controlling* the robot was perceived as being (see Figure 4, right), as well as how *difficult* both the task and interaction seemed and participants' *interest* in interacting with the depicted robot. In particular, women rated the robots more positively than did male participants as was observed in Experiment I. While this remains in contradition with prior work showing that men respond more positively towards robots than women (e.g., [15, 18, 20]), it nevertheless lends further support towards the methodological implication that gender is a relevant consideration for HRI studies.

Moreover, the results of the present investigation indicate that observatory perspectives of human-robot interactions are not substantially influenced by the robot's animacy. This suggests that simplistic

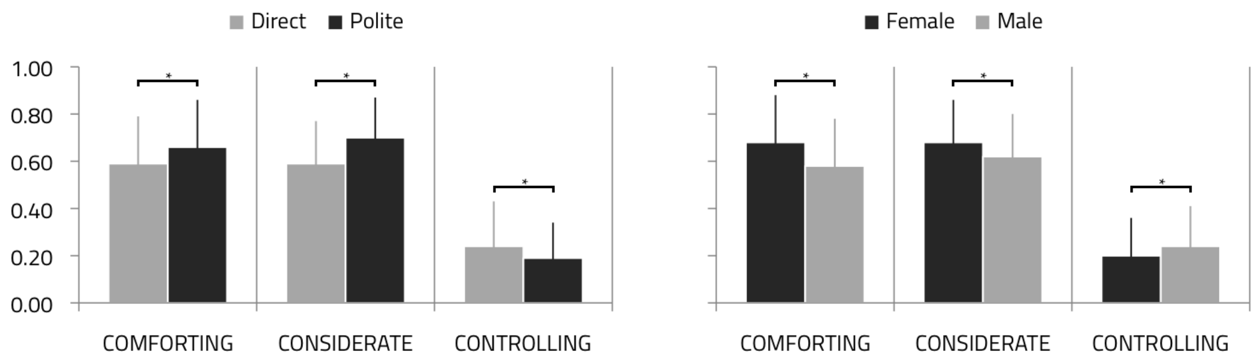

Figure 4: Main effects of *politeness* (left) and participant *gender* (right) on perceptions of the robot as *comforting*, *considerate*, and *controlling*. Dark bars emphasize factors yielding more positive outcomes (polite speech, female participants). All contrasts are significant.

depictions of human-robot interactions, such as in Experiment I, may suffice to investigate perceptions of certain robot behaviors (e.g., a robot's politeness, as perceived by female observers).

However, key differences in findings between the two experiments also underscore the necessity of considering perceptions in more realistic interaction scenarios. Specifically, unlike in Experiment I, Experiment II showed no interactions between any of the three IVs. For example, in the context of the more naturalistic interactions, the use of polite speech was effective at improving ratings regardless of the participant's gender. Whereas, in Experiment I, polite speech was only effective at improving *female participants'* ratings (while male participants of Experiment I were not receptive – the use of politeness modifiers, in the absence of the idling and attention sharing movements, did not improve ratings). This suggests that, while women appear to be particularly sensitive to verbal communication (as evidenced by their ratings across both Experiment I and the more naturalistic Experiment II), men may be more sensitive to *consistency* in verbal and nonverbal communicatory cues. Thus, the findings may imply a need for coherence between a robot's verbal and nonverbal communication (e.g., [10]).

In addition, the present experiment showed a slight influence of age on perceptions of *comfort* with the robot and how *controlling* it seemed (see Figure 3), whereas E1 showed no significant effects owing to participants' age. These effects are somewhat difficult to interpret, however, as it is unclear what aspects of the more realistic interaction would cause the standard university-aged participants (relative to the older adults) to here indicate less comfort with the robot and rate it as more controlling.

3.3.2 Limitations & Future Directions

Here we undertook further investigation of perceptions of robot politeness and potential modulatory factors. Our approach tested a few simple behaviors to assess the influence (or lack thereof) of a robot's movement. In particular, the presentation of human-robot interactions that were more naturalistic (i.e., mimic attention-sharing and idling behaviors exhibited in equivalent human-human interactions) compliments our previous study, which lacked the same degree of social realism. This contribution here is significant because it replicates the influences of politeness of both prior work and our own Experiment I. Further, it sheds light on how subject-based factors (i.e., age and gender) can yield more positive social evaluations. However, as with the previous study, our approach still has its limitations.

In particular, we explored here only a small subset of human-inspired movements. Thus, it is not possible to conclusively say that movement (of any kind) is effective for improving interactions or perceptions thereof. There are substantially more possibilities to try, such as gaze aversion (e.g., [1]) or gesturing (e.g., [11]) to name a few. To determine what extent certain types nonverbal communicatory mechanisms influence perceptions, future work might consider independently manipulating several types of movements, rather than the movement/no-movement meta comparison we made here.

4 GENERAL DISCUSSION

4.1 General Findings & Implications

As expected, Experiment I confirms prior indications that, at least in 3rd-person observation of pre-recorded human-robot interactions ([21, 22]), a robot's use of politeness modifiers in its speech is perceived more favorably relative to a robot that uses more direct speech (e.g., [14, 19, 21, 22]). This is reflected by participants ratings of the polite robot instructors as more *comforting* and *considerate*, and less *controlling* than the robots that were more direct. Moreover, the implications of politeness hold, even for a population that is highly diverse in terms of the socio-demographic factors of education, geographical location, age, and gender. Furthermore, we observed additional validation of the effects owing to a robot's politeness in Experiment II. Thus, consistent with prior indications ([21, 22]), the persistence of effects due to politeness – given the broader population sampling – demonstrate the benefit to using politeness modifiers when a robot communicates with natural language.

The results observed across the two studies further underscore an important methodological consideration – namely, gender – for evaluation of human-robot interactions. Specifically, we found a gender-based divide in the efficacy of the politeness manipulation in both experiments showing that a robot's use of politeness modifiers in its speech is most (and in Experiment I, only) effective for female participants. That is, here women rated polite robots significantly better than those that are more direct, and moreover, their ratings of polite robots are significantly higher than men's ratings of the same robots. Furthermore, the two studies suggest that men are sensitive to consistency in communicatory cues, and more importantly, they are not receptive to polite speech alone. These findings demonstrate the importance of considering gender – either as a systematic manipulation or as a covariate – in the analysis of human-robot interactions.

4.2 General Limitations & Future Work

Our approach to understanding perceptions of polite robots contributes a simple online task to assess the modulatory influences of various situational factors. We emphasize the benefit that the online forum serves for obtaining data with broad socio-demographics versus in-laboratory studies which are limited to smaller and more homogeneous participant populations. This lends the ability towards replicating previously indicated influences of politeness and understanding how such findings might transfer to the general population.

However, we wish to also underscore the limitations of this type of assessment. Despite the benefits to online studies, the results cannot be immediately applied to actual human-robot interactions involving co-located, direct participation, as the present work was conducted from a remote and observatory position (relative to the depicted interactions). Hence, whether (and if so, the extent to which) these findings generalize and apply to in-person, direct interactions with a co-located embodied agent motivates further investigation.

Further, we stress that these findings are preliminary and of limited weight. In particular, we note the small effect sizes observed across both studies. Between the two experiments, the effect sizes reached at most a medium qualification with the influence of politeness on perceptions of the robot as *considerate* ($\eta^2 = .11$ in the more naturalistic interaction scenario of Experiment II, and $\eta^2 = .06$ in Experiment I). Gender also showed an effect of close to a medium size on ratings of *comfort* ($\eta^2 = .06$). However, the size of other effects observed (e.g., due to age) is small ($\eta^2 \leq .03$). Thus, relative to other factors (e.g., the robot's appearance), the robot's politeness and the person's age/gender may be of little importance. While the present work yields implications for both the design of robotic agents and how to evaluate them, future work might consider how relevant gender and politeness are in other contexts or in contrast to other factors.

5 CONCLUSIONS

The primary aim of this research was to investigate whether previous results about human observers' preferences for polite robot speech over more direct speech in an robot instructor would hold for a wider participant demographic, which we were able to confirm. A secondary aim was to explore the modulatory influences of a person's age and gender on perceptions of the robot. Here we obtained several new and important gender effects that hint at a complex interplay of the interaction observers' gender with the observed robot's behavior, which warrants further investigation to elucidate the causal mechanisms responsible for the gender-based differences. Further, owing to a limitation of the design of our first experiment, we explored peoples' perceptions given a more realistic interaction scenario which additionally confirmed the influence of both politeness and gender. These findings are particularly important for the design of future autonomous agents, robotic or virtual, because their success could significantly depend on their ability to adapt, such as to gender-specific expectations of their interactants.

REFERENCES

[1] Sean Andrist, Xiang Zhi Tan, Michael Gleicher, and Bilge Mutlu, 'Conversational gaze aversion for humanlike robots', in *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*, pp. 25–32, (2014).

[2] Christoph Bartneck, Tomohiro Suzuki, Takayuki Kanda, and Tatsuya Nomura, 'The influence of peoples culture and prior experiences with aibo on their attitude towards robots', *AI & Society*, **21**(1), 217–230, (2007).

[3] Charles R Crowell, Michael Villano, Matthias Scheutz, and P. Schermerhorn, 'Gendered voice and robot entities: perceptions and reactions of male and female subjects', in *Proceedings of the IEEE International Conference on Intelligent Robots and Systems*, pp. 3735–3741, (2009).

[4] Juan Fasola and Maja J. Mataric, 'A socially assistive robot exercise coach for the elderly', *Journal of Human-Robot Interaction*, **2**(2), 3–32, (2013).

[5] Priska Flandorfer, 'Population ageing and socially assistive robots for elderly persons: the importance of sociodemographic factors for user acceptance', *International Journal of Population Research*, (2012).

[6] ME Foster, A Gaschler, M Guiliani, A Isard, M Pateraki, and RPA Petrick, 'Two people walk into a bar: Dynamic multi-party social interaction with a robot agent', in *International Conference on Multimodal Interaction (ICMI)*, pp. 3–10, (2012).

[7] Bruce Fraser, 'Conversational mitigation', *Journal of Pragmatics*, **4**, 341–350, (1980).

[8] Kerstin Sophie Haring, David Silvera-Tawil, Yoshio Matsumoto, Mari Velonaki, and Katsumi Watanabe, 'Perception of an android robot in japan and australia: A cross-cultural comparison', in *Social Robotics*, 166–175, (2014).

[9] Lewis Hassell and Margaret Christensen, 'Indirect speech acts and their use in three channels of communication', in *Communication Modeling*, p. 9, (1996).

[10] J Hodgins, S Jörg, C O'Sullivan, SI Park, and M Mahler, 'The saliency of anomalies in animated human characters', *ACM Transactions on Applied Perception (TAP)*, **7**(4), 1–14, (2010).

[11] Heeyoung Kim, Sonya S Kwak, and Myungsuk Kim, 'Personality design of sociable robots by control of gesture design factors', in *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication*, pp. 494–499, (2008).

[12] VA Kulyukin, 'On natural language dialogue with assistive robots', in *HRI*, pp. 164–171, (2006).

[13] I Han Kuo, Joel Marcus Rabindran, Elizabeth Broadbent, Yong In Lee, Ngaire Kerse, RMQ Stafford, and Bruce A MacDonald, 'Age and gender factors in user acceptance of healthcare robots', in *Proceedings of the IEEE International Conference on Robot and Human Interactive Communication*, pp. 214–219, (2009).

[14] Min Kyung Lee, Sara Kiesler, Jodi Forlizzi, Siddhartha Srinivasa, and Paul Rybski, 'Gracefully mitigating breakdowns in robotic services', in *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*, pp. 203–210, (2010).

[15] Bilge Mutlu, Steven Osman, Jodi Forlizzi, Jessica Hodgins, and Sara Kiesler, 'Task structure and user attributes as elements of human-robot interaction design', in *Proceedings of the IEEE International Conference on Robot and Human Interactive Communication*, pp. 74–79, (2006).

[16] Tatsuya Nomura and Kazuma Saeki, 'Effects of polite behaviors expressed by robots: A psychological experiment in japan', *International Journal of Synthetic Emotions*, **1**(2), 38–52, (2010).

[17] Tatsuya Nomura and Satoru Takagi, 'Exploring effects of educational backgrounds and gender in human-robot interaction', in *Proceedings of the International Conference on User Science and Engineering*, pp. 24–29, (2011).

[18] Natalia Reich and Friederike Eyssel, 'Attitudes towards service robots in domestic environments: The role of personality characteristics, individual interests, and demographic variables', *Journal of Behavioral Robotics*, **4**(2), 123–130, (2013).

[19] Maha Salem, Micheline Ziadee, and Majd Sakr, 'Marhaba, how may I help you?: Effects of politeness and culture on robot acceptance and anthropomorphization', in *Proceedings of the ACM/IEEE International Conference on Human-robot Interaction*, pp. 74–81, (2014).

[20] Paul Schermerhorn, Matthias Scheutz, and Charles R. Crowell, 'Robot social presence and gender: Do females view robots differently than males?', in *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*, pp. 263–270, (2008).

[21] Megan Strait, Cody Canning, and Matthias Scheutz, 'Let me tell you! Investigating the effects of robot communication strategies in advice-giving situations based on robot appearance, interaction modality, and distance', in *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*, pp. 479–486, (2014).

[22] Cristen Torry, Susan R. Fussell, and Sara Kiesler, 'How a robot should give advice', in *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*, pp. 275–282, (2013).