Asking for Help from a Gendered Robot

Emma Alexander, Caroline Bank, Jie Jessica Yang, Bradley Hayes, Brian Scassellati
Department of Computer Science, Yale University
51 Prospect St, New Haven, CT 06511
{caroline.bank,emma.alexander,jiejessica.yang,bradley.h.hayes,brian.scassellati}@yale.edu

Abstract

This project investigates the effects of gender in a human-robot collaboration interaction. In the experiment, participants completed four Sudoku-like puzzles with a robot from which they could verbally elicit help. The robot was given the gendered characteristics of a gendered computer generated voice and either the name Charlotte (female condition) or Charley (male condition). Contrary to expectations from psychology, male participants asked the robot for help more frequently regardless of its assigned gender. Participants of both genders reported feeling more comfortable with a robot assigned the other gender and preferred the male robot's help. Findings indicate that gender effects can be generated in human-robot collaboration through relatively unobtrusive gendering methods and that they may not align with predictions from psychology.

Background

Gender effects are present and well-studied in human-human interaction and collaboration. When presented with the description of a potential helper, men have been seen to be more likely to seek help from a woman than from a man (Nadler, Maler, & Friedman, 1984). On the other hand, mixed-gender groups working on collaborative computer-based tasks exhibit less collaborative behavior (Collazos, Guerrero, Llana, & Oetzel, 2002) and lower performance (Underwood, McCaffrey, & Underwood, 1990) than same-sex groups in identical circumstances. Women were also more likely to seek help in general (Nadler et al., 1984). Women and men also show differences in asking for help from attractive versus unattractive members of the same or different gender (Nadler, Shapira, & Ben-Itzhak, 1982). These effects also carry over to perceived helpfulness: Menzel and Carrell (1999) found a strongly mediated but nonetheless significant effect on perceived learning where students believed they learned more when taught by a professor of the same gender.

Existing literature suggests that helpfulness is not the only quality that people perceive through a lens of gender. In 1968, Rosenkrantz et al found that over 75% of participants of both genders associated a number of attributes to one gender or the other. In particular, participants associated competence, dominance, independence, and logic with males, and emotion, subjectiveness, submission, and tact with females (1968). These stereotypes were reaffirmed in a survey by Hosoda and Stone (2000). Biernat and Kobrynowicz found that these biases can affect other behavior, as subjects that were asked to evaluate candidates in a mock job interview setting rated woman and minorities against a lower standard for baseline competence but more stringent standards for high levels of competence (1997; 2010). Even gendered voices with no other stimuli evoke different responses in the brain (Lattner, Meyer, & Friederici, 2005; Decasper & Prescott, 1984).

After interacting with a robot, people will often bring up comments on robot gender unprompted, demonstrating that its strength as a social cue is still highly relevant (Carpenter et al., 2009). In another study, participants used more complex language when describing the stereotypically female domain of relationships to a female robot than a male robot (Powers et al., 2005). Men have also been shown to be more willing to make donations when asked by a female robot (Siegel, Breazeal, & Norton, 2009). This study also demonstrated that after the interaction, participants saw robots of the other gender as more credible, trustworthy, and engaging, indicating an other-gender preference in the case of a robot asking a human for a help. It has been found that in particular situations, people prefer interaction with an other-gender robot (Park, Kim, & del Pobil, 2011) and will consider a robot displaying other-gender characteristics more attractive, interesting, and trustworthy than one displaying same-gender characteristics (Siegel et al., 2009). In many studies, gender was conferred to the robot using only a single, relatively subtle cue: a gendered name and introduction (Nomura & Takagi, 2011), voice modulation (Siegel et al., 2009), a combination of the two (Crowelly, Villanoy, Scheutzz, & Schermerhornz, 2009), or voice modulation accompanied by variation in lip color (Powers et al., 2005). These results indicate that even with limited gender cues behavioral responses can be elicited from humans by robotic stimulus. This study explores the shaping of perceptions within the specific case of a human seeking help from a gendered robot collaborator.

Methods

To study gender effects on human-robot collaboration we designed a 2x2 study on robot and participant gender. Robot gender was assigned with a name/pronoun cue and a voice cue. Past studies have shown that these cues alone are sufficiently strong to elicit a gender response. Changes in appearance were avoided as they could confound results. We picked a puzzle-based task because it provided a way for the robot to contribute to the participant's success while allowing the variety of interaction to be limited, ensuring that participants shared a standard experience. The puzzle chosen was sudoku, with the numbers 1-9 replaced with the letters A-I so that both analytical and verbal skills would be indicated and the robot perception would not be affected by stereotypes of numerical intelligence. Participants were given four puzzles in order of increasing difficulty to control for differences in their previous experience with the activity. Each puzzle was in-

	1	2	3	4	5	6	7	8	9
A		I	C	A	H			G	
B	A					B		C	H
C	H				C	G	I	F	A
D		A	I	C	D	H	G		F
E	C	H			B		A		I
F	G	F		E	I	A			C
G	B			H	A	D		I	
H			H			I		A	G
I	I		A		G	C	H		

(a) Sample puzzle, unsolved

	1	2	3	4	5	6	7	8	9
A	F	I	C	A	H	E	B	G	D
B	A	D	G	I	F	B	E	C	H
C	H	B	E	D	C	G	I	F	A
D	E	A	I	C	D	H	G	B	F
E	C	H	D	G	B	F	A	E	I
F	G	F	B	E	I	A	D	H	C
G	B	G	F	H	A	D	C	I	E
H	D	C	H	B	E	I	F	A	G
I	I	E	A	F	G	C	H	D	B

(b) Sample puzzle, solved

Figure 1: An example of the puzzles participants were to solve with the robot.

tentionally allotted a shorter time period than was needed for completion to encourage interaction with the robot. During each puzzle, researchers tallied the number of help requests from each participant. Additionally, participants were asked to rate the robot on a number of qualities (trustworthiness, intelligence, etc.) to further characterize the interaction.

Hypotheses

- **H1** Women would ask more questions of the robot than men, regardless of robot gender.
- **H2** Participants of both genders would ask more questions of a robot assigned the opposite gender.

Procedure

Participants were presented with an entrance survey and a warm-up puzzle as an introduction to the task. They were told by the experimenter that the purpose of the study was to aid development of robots for use in human-robot collaboration in puzzle solving. They were then introduced to the robot and told that it had been developed to collaborate with humans on the task. Participants were informed that the robot could not see, but that it could speak and hear and that the experimenter would be "downloading" the puzzle to the robot.

The task used in this experiment was a set of Sudoku puzzles over the letters A through I instead of the conventional numbers 1 through 9. The puzzle is a 9x9 grid into which each of the letters must be placed so that each letter appears in every row, column, and 3x3 square once. The grid's rows were labelled with the letters A through I and its columns were numbered 1 through 9. This system was explained to the participants as a way to communicate with the robot, for example referring to the upper-left cell in a puzzle as "square A1."

Participants each completed four puzzles with the robot. Each trial (puzzle) was limited to five minutes, and participants were told to begin and end work upon verbal cues from the robot. The starting cue was "*Hi, my name is Charley/Charlotte. Let's work together*" for the first puzzle and "*I have received the puzzle. Let's get started*" for the remaining three puzzles. The robot was remotely operated by the experimenter based on audio monitoring of the interaction. For responding to participant queries, the experimenter was able to select from the following phrases:

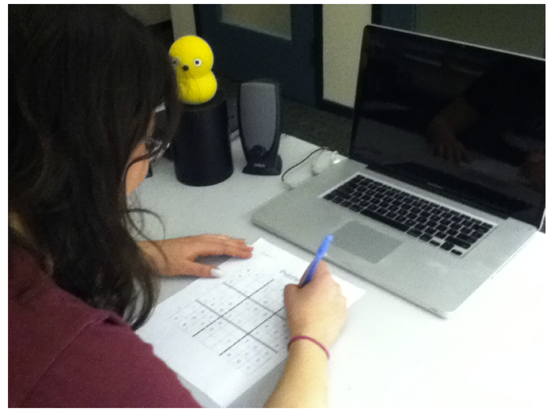
- I think so
- I don't think so
- There is/are [number] minutes remaining
- I think there is a [letter] in square [coordinates]

If the participant asked a question that could not be answered with one of these phrases, the robot would respond with "*I don't know how to answer that*". The robot also automatically informed the participant when one minute remained in each trial, and at the end of each trial with the phrase "*time is up*". Before speaking, the robot moved its body down, tilted forward, and paused for two seconds to emulate thinking. At the end of each trial, the experimenter returned and administered a brief survey.

Participants

The 48 participants used in this study were members of the Yale community (24 female and 24 male). Data were collected over a 12-week period. Participants were compensated for their time with entry into a gift card lottery.

(a) Keepon Pro robot



(b) Experimental setup

Figure 2: The robot platform and experimental environment used in our study.

Robot

The robot used in this study was a Keepon, seen in Figure 2a. Keepon is a small, yellow robot that has previously been used to study human-robot interaction (Kozima, Michalowski, & Nakagawa, 2009). The robot sits on a stationary platform and has four degrees of freedom allowing it to rotate on the platform, tilt backwards and forwards, lean sideways, and bounce up and down. It has cameras in its eyes and a microphone in its nose.

Conditions

Each participant was randomly assigned to either the female or male robot experimental group upon arrival. This, plus participant gender, created four experimental groups: male-participant/female-robot (M/F), female-participant/female-robot (F/F), male-participant/male-robot (M/M), and female-participant/male-robot (F/M). The experimenter referred to the robot as "Charlotte" in the female robot conditions and as "Charley" in the male robot conditions. This name change was reflected in the experimental surveys, which also referred to the robot by the appropriate gendered pronoun. The gendering was reinforced by the choice of the robot's computer generated voice: Mac OS X's "Vicki" for the female robot and "Alex" for the male robot.

Measures

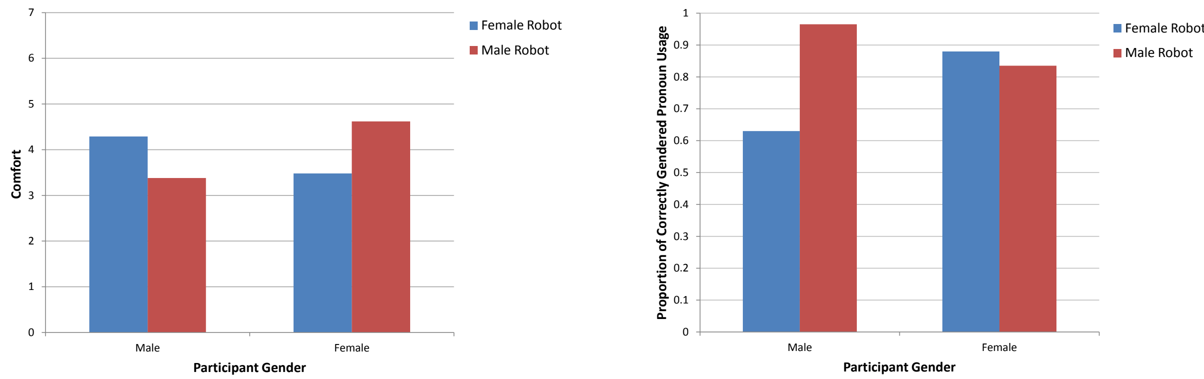
The experimenter operating the robot recorded the number of successful and failed interactions between the participants and the robot. Failed interactions were defined as any incidents of participants asking questions that necessitated the "*I don't know how to answer that*" response, with successful interactions defined as any other incidents of participants asking questions. The puzzles were also scored and recorded. The surveys completed after the first, second, and third puzzles contained five 7-point Likert-scale questions on the robot's perceived intelligence, competence, helpfulness, trustworthiness, and likability. The fourth survey contained seventeen 7-point Likert-scale questions in addition to several free-response questions.

Participant Gender	Robot Gender	Mean	SD
F	F	-2.958	1.994
	M	-4.25	1.994
M	F	-5.333	1.994
	M	1.083	1.994

Table 1: Difference between participants' estimated number of requests and actual number of requests on the first puzzle

Results

Data were collected from 48 participants, 12 in each condition. Participant scores were calculated by totaling the number of correctly placed letters in each puzzle at the end of the trial time. There was no significant effect of either participant gender ($F(1,44) = .023$, $p = 0.881$) or robot gender ($F(1,44) = 1.479$, $p = .230$) on this measure of task performance. A two-way analysis of variance found a borderline significant effect of participant gender on the number of successful interactions ($F(1,44) = 3.263$, $p = .078$, Female participants: $M = 58.125$, $sd = 5.725$, Male participants: $M = 72.750$, $sd = 5.725$) and a significant effect of gender on the number of failed interactions($F(1,43) = 6.836$, $p=.012$, Female participants: $M = .750$, $sd = .304$, Male participants: $M = 1.875$, $sd = .304$) initiated with the robot. In the exit survey, participants were asked for an estimate of how many times they asked the robot for help during each puzzle. There was a borderline significant participant gender-robot gender interaction on the difference between this measure and the actual number of requests for the first puzzle. A summary of this result can be found in Table 1. The participant gender effect on number of interactions may have resulted in a different experience for male participants versus female participants. To account for this



(a) Adjusted mean ratings of comfort.

(b) Adjusted mean proportion of correctly-gendered pronoun use.

Figure 3: The above are plots of adjusted means calculated by performing analyses of covariance using total successful interactions and total failed interactions as covariates

difference, all further analysis was performed controlling for both measures. An analysis of covariance with robot gender and participant gender and between-subjects with the number of failed interactions and number of successful interactions as covariates found a significant interaction between participant and robot gender interaction on ratings of comfort with the robot ($F(1,42) = 6.590$, $p = .014$). Mean ratings of comfort by both groups of participants were higher for the robot of the opposite gender - a plot of adjusted means can be found in Figure 3a. A further analysis of covariance found a that participants rated their preference for working with the robot over a human partner as higher for the male robot than for the remale robot ($F(1,42) = 12.347$, $p = .001$, Female robot: adjusted mean = 3.657, sd = .373, Male robot: adjusted mean = 5.412, sd = .373). In the exit survey, participants were asked to give answers to five free response questions, four of which pertained to their feelings towards the robot. These responses were scored for gendered pronoun usage. The generic pronouns *it* and *its* were scored as 'non-gendered', *she* and *her* were scored as 'correctly-gendered' for participants who interacted with Charlotte and 'cross-gendered' for participants who interacted with Charley. *He* and *him* were scored as 'cross-gendered' for participants who interacted with Charlotte and 'correctly-gendered' for participants who interacted with Charley. Only a single participant (in the 'M/F' condition) used cross-gendered pronouns. However, there was a borderline significant interaction of robot gender and participant gender on the proportion of total pronouns used that were 'correctly-gendered'. A plot of the adjusted means can be found in Figure 3b.

Discussion

In this experiment, participants were asked to engage in puzzle solving with a characteristically gendered robot. It remains unclear whether the results found were driven by a gender bias or by reactions to the specific ways the robot's gender was manipulated. It is possible, for instance, that the differences found were differing reactions to the computer-generated voices chosen. Furthermore, previous research has found that people tend to identify robots as either genderless or male (Shin & Kim, 2007). It must be noted that age and culture differences may make Shin's result inapplicable to the demographic targeted here since his participants were Korean schoolchildren rather than American young adults. However, in the absence of a comparable study performed at an American university this result is the closest we have to a characterization of robot gender assumption. Then we expect that if a participant identifying a robot as male was confronted by the female voice and naming given to Charlotte, it is possible that the dissonance between their perception of the robot and its gendered characteristics may have made the robot a less desirable partner. This effect may have been particularly present for male participants who worked with the female robot, who on average referred to the female robot as 'it' more frequently in proportion to gendered pronouns than men with a male robot or women with either robot gender.

Contrary to our first hypothesis (**H1**), male participants initiated more interactions with the robot than female participants. This stands in contrast with previous research in psychology that has found men less willing to ask humans for help. Results from psychology support the notion that men are more comfortable with novel technology than women are, an effect that has been confirmed in the domain of human-robot interaction (Carpenter et al., 2009). This may have influenced their willingness to ask the robot for assistance. At the same time, female participants rated themselves as liking the robot of both genders at a higher level than male participants at a borderline significant level($F(1,42) = 3.466$, $p = .070$, Female participants: $M = 5.726$, $sd = .280$, Male participants: $M = 4.941$, $sd = .280$). This may have been because while male participants felt more comfortable asking for assistance, this asking inspired negative feelings towards the robot. It is also possible that within the framework of the experiment, the interactions were not perceived as asking for help. One participant referred to the robot as an "answer key" either because of the nature of the questions asked or a lack of agency attributed to the robot. If participants saw themselves as using the robot rather than interacting with it, we would expect the results from the human-robot interaction literature to dominate the effects predicted from collaboration psychology.

Contrary to our second hypothesis (**H2**), we did not find that participants were more likely to ask for help from a robot assigned the opposite gender. However, participants interacting with a robot of the opposite gender on average rated the robot as making them feel more comfortable. This comfort may have stemmed from a number of different aspects of the interaction. Given that the participants were prompted to ask the robot for help, this result might reflect their comfort in seeking assistance from the robot. This would support previous research that found men to be more willing to ask for help from women and vice versa, and with studies that show more positive interactions with robots of the other gender (Siegel et al., 2009; Park et al., 2011; Powers et al., 2005). It is also possible that participants were more comfortable with the different computer voice, or simply from interacting with a different-gendered robot.

Both of our hypotheses, based on previous study of human-human collaboration, were incorrect. Our results were, however, supported by studies of non-collaborative human-robot interaction, which suggests that the type of the interaction is less important than the type of interactive partner. On the other hand, psychological effects may have played a role in the perception of the interaction. This could explain why men underestimated the number of questions they had asked the female robot and overestimated the number of requests to the male robot. This means that within the domain of each interaction, psychology literature cannot predictably be applied to gendered interactions with robots and when multiple social factors are involved, as in this experiment, the behaviors and perceptions of participants must be measured rather than inferred from previous studies. Further research is needed to elucidate and confirm the gender characteristics that evoke different reactions, and the social dynamics that underlie those differences.

To complicate this picture, participants rated the male robot more highly than the female robot for their agreement to the statement "I would prefer working with Charlotte/Charley over a human partner for this task." However, it is unclear in what way the female robot was perceived as comparatively less desirable. There were no significant effects of robot gender on the related ratings of liking and the robot's ability to solve the puzzle. Further study is required to determine whether this preference was due to the robot's perceived social abilities, task-solving abilities, or other factors.

Conclusion

In this experiment we presented participants with a series of puzzle-based tasks and a robot partner to assist them. This robot was assigned a gender through seemingly unobtrusive cues: name and voice modulation. Contrary to expectations based on psychological literature, male participants engaged in more help-seeking interactions with the robot than did female participants, perhaps because of the effectiveness of the help, which has been shown to influence men's technology usage more than that of women(Venkatesh & Morris, 2000). Also in contrast with psychology literature, we did not find that participants asked more questions of a robot of the other gender, but participants did report that they were more comfortable with the robot in that case. Participants of both gender reported a preference for working with the male robot, though it is unclear if this is an effect of gender or the specifics of how the robots were gendered. Overall, these results indicate that human-robot interactions can be significantly affected by fairly subtle gender cues, and that these effects may not be reliably predictable from studies of human-human interactions. Because our results more closely mirrored those from other kinds of human-robot interactions than those of human collaborations, we suspect that a task with expanded opportunity for collaboration, which was requested by many participants, may foster interactions which more closely match psychological results. However, the required range of possible interactions would make it difficult to ensure standardization or even similarity of participant experience.

Acknowledgments

This work is supported by NSF grants 1139078 and 1117801, and Office of Naval Research grant #N00014-12-1-0822. We thank the anonymous reviewers for their helpful comments.

References

Biernat, M., Fuegen, K., & Kobrynowicz, D. (2010). Shifting standards and the evaluation of competence: Complexity in gender-based judgment and decision making. *Personality and Social Psychology Bulletin*, *36*(7), 855-868.

Biernat, M., & Kobrynowicz, D. (1997). Gender- and race-based standards of competence: Lower minimum standards but higher ability standards for devalued groups. *Journal of Personal and Social Psychology*, *72*(3), 544-557.

Carpenter, J., Davis, J. M., Erwin-Stewart, N., Lee, T. R., Bransford, J. D., & Vye, N. (2009). Gender representation and humanoid robots designed for domestic use. *International Journal of Social Robotics*, *1*(3), 261–265.

Collazos, C., Guerrero, L. A., Llana, M., & Oetzel, J. (2002). Gender: An influence factor in the collaborative work process. *Proceedings of the 4th International Conference on New Educational Environment (ICNEE 2002)*, 7-10.

Crowelly, C., Villanoy, M., Scheutzz, M., & Schermerhornz, P. (2009). Gendered voice and robot entities: perceptions and reactions of male and female subjects. In *Intelligent robots and systems, 2009. iros 2009. ieee/rsj international conference on* (pp. 3735–3741).

Decasper, A. J., & Prescott, P. A. (1984). Human newborns' perception of male voices: Preference, discrimination, and reinforcing value. *Developmental Psychobiology*, *17*(5), 481-491.

Hosoda, M., & Stone, D. L. (2000). Current gender stereotypes and their evaluative content. *Perceptual and Motor Skills*, *90*, 1283-1294.

Kozima, H., Michalowski, M. P., & Nakagawa, C. (2009). Keepon. *International Journal of Social Robotics*, *1*(1), 3–18.

Lattner, S., Meyer, M. E., & Friederici, A. D. (2005). Voice perception: Sex, pitch, and the right hemisphere. *Human Brain Mapping*, *24*(1), 11-20.

Menzel, K. E., & Carrell, L. J. (1999). The impact of gender and immediacy on willingness to talk and perceived learning. *Communication Education*, *48*(1), 31-40.

Nadler, A., Maler, S., & Friedman, A. (1984). Effects of helper's sex, subjects' androgyny, and self-evaluation on males' and female's willingness to seek and receive help. *Sex Roles*, *10*(5-6), 327-339.

Nadler, A., Shapira, R., & Ben-Itzhak, S. (1982). Good looks may help: Effects of helper's physical attractiveness and sex of helper on males' and females' help-seeking behavior. *Journal of Personality and Social Psychology*, *42*(1), 90-99.

Nomura, T., & Takagi, S. (2011). Exploring effects of educational backgrounds and gender in human-robot interaction. In *User science and engineering (i-user), 2011 international conference on* (p. 24-29).

Park, E., Kim, K. J., & del Pobil, A. P. (2011, May). The effects of robot's body gesture and gender in human-robot interaction. In *Proc. 2011 iasted international conference on internet and multimedia systems and applications (ieee/hci).* Washington, DC, USA.

Powers, A., Kramer, A., Lim, S., Kuo, J., lai Lee, S., & Kiesler, S. (2005, aug.). Eliciting information from people with a gendered humanoid robot. In *Robot and human interactive communication, 2005. roman 2005. ieee international workshop on* (p. 158 - 163). doi: 10.1109/ROMAN.2005.1513773

Rosenkrantz, P., Vogel, S., Bee, H., Broverman, I., & Broverman, D. M. (1968). Sex-role stereotypes and self-concepts in college students. *Journal of Consulting and Clinical Psychology*, *32*(3), 287-295.

Shin, N., & Kim, S. (2007, aug.). Learning about, from, and with robots: Students' perspectives. In *Robot and human interactive communication, 2007. ro-man 2007. the 16th ieee international symposium on* (p. 1040 -1045). doi: 10.1109/ROMAN.2007.4415235

Siegel, M., Breazeal, C., & Norton, M. (2009, oct.). Persuasive robotics: The influence of robot gender on human behavior. In *Intelligent robots and systems, 2009. iros 2009. ieee/rsj international conference on* (p. 2563 -2568). doi: 10.1109/IROS.2009.5354116

Underwood, G., McCaffrey, M., & Underwood, J. (1990). Gender differences in a cooperative computer-based language task. *Educational Research*, *32*(1), 44-49.

Venkatesh, V., & Morris, M. G. (2000). Why don't men ever stop to ask for directions? gender, social influence, and their role in technology acceptance and usage behavior. *MIS Quarterly*, *24*(1), 115-139.