International Journal of Gender, Science and Technology
http://genderandset.open.ac.uk

It's What You Call It: Gendered Framing and Women's and Men's Interest in a Robotics Instruction Task

Sarah Morton, Julie A Kmec, Matthew E Taylor

Washington State University, USA

ABSTRACT

While women played a pivotal role at the onset of computer programming, technology fields, including robotics, are currently gendered. Prior research fails to consider how the context of robotics activities might impact men's and women's perceptions of robotics tasks. Our study, a survey experiment of 132 men and women, seeks to understand 1) whether women and men have different interest in robotics instruction tasks, 2) whether describing a robotics instruction task scenario as feminine (teaching the robot), masculine (programming the robot), or gender neutral (training the robot) relate to women's and men's interest in robotics instruction tasks, and 3) whether perceptions that instructing robots to perform tasks is good for society or requires natural talent impacts women's interest in the robotics task more strongly than men's interest. Contrary to our hypotheses, we find that women are more interested in most of the robotics instruction tasks than are men, and that framing the robotics instruction task scenarios as feminine produces *worse* outcomes for women in some of the robotics tasks. We also find that both men and women who think instructing robots to perform tasks is beneficial to society are significantly more likely to have interest in the robotics tasks than those who do not hold these views. We provide possible explanations for these findings, as well as theoretical and practical implications.

KEYWORDS

gender, STEM education, human-robot interaction, technology

It's What You Call It: Gendered Framing and Women's and Men's Interest in a Robotics Instruction Task

INTRODUCTION

Technology fields are currently gendered in the U.S.; that is, women and men place themselves in relation to technology based on masculine and feminine norms (Witt & Hofmeister, 2015; Davison & Argyriou, 2016). The growing field of human robot interaction, or HRI, is an ideal place to investigate the gendering of technology. This is because as robots become increasingly common, the currently dominant paradigm of robots acting in isolation (e.g., a robot on an automotive factory production line) will necessarily be replaced by robots working in close proximity with humans. One common type of HRI will be that of *instruction*. Of the infinite number of different tasks that robots can be programmed to accomplish, only a limited number can be included in a robot's default package, so humans will have to train robots to adapt and learn in the specific environment in which they are deployed. In many cases, robots will not be able to learn autonomously (learning is ill-formed if there is no well-defined objective) but instead, in conjunction with humans. Humans will help robots learn through multiple modalities, including by direct programming, demonstrations of correct behavior (Chernova & Thomaz, 2014), evaluative feedback (Knox & Stone, 2008; Loftin et al., 2016), direct rewards (Thomaz & Breazeal, 2006), or natural language advice (Kuhlmann et al., 2004). To date, little research explores whether people *want* to instruct robots, what types of robot instruction methodologies are most useful in different settings, or how to frame robot instruction so that it is easiest for the human.

Nor have researchers asked these questions with an eye toward the way women and men may differently interact with the instruction of robots. Indeed, research on gender and HRI is in a nascent stage; most HRI literature that also explores gender focuses on perceptions of robots when they are gendered as male or female. For example, in a set of experimental studies that have "gendered" robots as male or female using facial cues and voices, researchers have found that participants perceived the "male" robots as more competent at stereotypically masculine tasks (e.g., repairing technical devices) than "female" robots (Nass, Moon & Green, 1997; Eyssel & Hegel, 2012). However, another study found that subjects rated "female" and "male" robots as equally competent at performing a stereotypically feminine task, but that the "female" robot was considered to be more competent than the "male" robot engaged in a stereotypically masculine task (Kuchenbrandt et al., 2014).

To date, this emerging field of study at the intersection of gender and HRI has failed to consider how the context of robotics activities might impact gendered perceptions of completing technological tasks. Our study contributes to this developing area by paying particular attention to how the gendered framing (or description) of a robotics instruction task scenario may be related to men's and women's perceptions of instructing robots to perform different tasks. Specifically, we present data from an original survey experiment designed to examine how men and women perceive several robotics tasks in a robotics instruction scenario that is experimentally manipulated to be either masculine, feminine, or neutral. We address three research questions: (1) To what extent are there gender differences in interest in robotics instruction tasks? (2) Are women's and men's interests in robotics instruction tasks dependent upon whether the robotics instruction task scenario is framed as feminine, masculine, or gender neutral? And (3) Do perceptions that the instruction of robots to perform tasks are either good for society or require natural talent differently impact women's and men's interest in the robotics task?

A robotics instruction scenario is both interesting and important for studying gender and HRI for theoretical, methodological, and policy reasons. First, given the broad reach of gender and gender essentialist assumptions of women's and men's capacities (see Ridgeway, 2011), knowing if and how the gendered framing of a robotics instruction task scenario as masculine or feminine is crucial to know in order to develop best practices for increasing women's participation in the field of computer science and related STEM fields. In the United States, women currently earn about 18 percent of the bachelor's degrees in computer science, a percentage that has remained relatively stagnant over the last decade (National Science Foundation, 2016). Second, from a methodological standpoint, a robotics instruction task scenario can be easily manipulated as masculine, feminine, or gender neutral, thereby making robotics instruction task scenarios ideal to study the salience of gender. Third, as robotics technology develops, humans will have to interact with robots more on the job, so understanding if and how human-robot interactions differ for women and men is necessary for engaging the largest possible workforce. As such, our research has practical implications for the workplace (Mutlu & Forlizzi, 2008).

GENDER and STEM

Where and how men's greater interest in computer science and technology emerges is up for debate, but evidence suggests it is not because of academic achievement differences between men and women (Riegle-Crumb, King, Grodsky, & Muller, 2012). Indeed, in the United States, women and men perform similarly in subjects that form the basis for computer science. That is, women and men perform roughly the same in development of early number concepts and earn higher grades in mathematics courses than boys through the end of high school; some have even found no gender difference in math achievement between girls and boys (see Else-Quest, Hyde & Lin,n 2010).

We draw on the broader body of research on gender and technology stereotypes, specifically on gender stereotypes in STEM (Science, Technology, Engineering, and Math) education, to frame our study. While it is beyond the scope of this manuscript to review this body of literature, we focus mainly on findings related to computer science and technology. It is important to note that we focus on findings from the United States and other Western countries, as women's representation in computer science, engineering, and technology in these countries is substantially lower than in many Asian, Middle Eastern, and African countries (Othman & Latih, 2006; Galpin, 2002; Lagesen, 2008).

First, a large body of literature has examined men's and women's interests in STEM fields in Western countries. From this literature, we know that men are more interested in computer science and technology degrees or taking courses in these fields than are women (Appianing & Van Eck, 2015). Furthermore, those who are interested in computer science, whether they are men or women, have been found to have a high computer self-efficacy, meaning they are comfortable with computers and confident in their ability to learn computing skills (Beyer, 2014).

Second, research has concluded that gender stereotyping in computer science and related disciplines ultimately discourages women from entering or remaining in these fields. At the macro level, countries with a higher representation of women in STEM at the tertiary level (community college and above) have lower gender-science stereotypes that associate men with science over women (Miller, Eagly & Linn, 2015). Interestingly, this finding held even if overall gender equity in a country was high, but the representation of women in STEM was low, which is the case in the United States and many other developed, Western countries. At the more micro-level, Wynn and Correll (2018) recently found that at technology company recruitment events, recruiters (who were almost all men) often resorted to behaviors that made women feel excluded, eventually deterring women from accepting or even applying for tech jobs. For example, recruiters engaged in gendered speech, "fraternity-like" banter, referred to highly technical aspects of the job (that only those with very high levels of skill could follow, even when the skill level was not required of the jobs they were recruiting for), and frequently discussed predominantly male cultural references.

Given men's greater interest in computer science and technology compared to women's, we hypothesize the following:

Hypothesis 1: Men's interest in the robotics instruction task will be greater than women's interest.

GENDERED FRAMING OF TASKS AND STEM EDUCATION

Another likely cause of women's lower interest in computer science, and STEM fields more generally, is the social context surrounding these fields of study. That is, researchers have found compelling evidence to suggest that social factors, including the framing of science environments, impact women's STEM interest. Several experiments have revealed that the framing of science environments influences men's and women's different interest in computer science. In one study, researchers manipulated the objects in a computer science classroom to be either stereotypical of computer scientists (e.g., Star Trek posters and video games) or not stereotypical of computer scientists (e.g., nature posters and phone books) and found that the non-stereotypical classroom was enough to bring female undergraduates' interest in computer science up to their male counterparts, bridging the gender gap in computer science interest (Cheryan et al., 2009). Another study drawing on two similar experiments found that women's lower interest in computer science courses may stem from stereotypes in computer science (Master, Cheryan, & Meltzoff, 2016). Not only did they find that girls presented with the non-stereotypical computer science classroom were more interested in computer science than girls presented with the stereotypical classroom, but that when the stereotypes in the classroom were evident, girls had a lower sense of belonging than boys, and this lower sense of belonging predicted girls' lower interest in computer science. In other words, the difference in computer science between girls and boys was mediated by their lower sense of belonging in a computer science course. Drawing on this body of literature, we hypothesize the following:

Hypothesis 2: Framing a robotics instruction task scenario using "feminine" or "gender neutral" language compared to "masculine" language (e.g., teaching, training, or programming a robot) will have a more positive association with women's interest in the task scenario than with men's interest.

Another related body of research shows how the context and framing of STEM education might be related to women's representation in STEM fields. In general, women have expressed interest in doing research that impacts society (Smith-Doerr et al., 2016) and women leave STEM fields in part because of the inability of university classrooms to make STEM education accessible or align with their goals to contribute to society (Espinosa, 2011). Reporting in the *New York Times*, Nilsson (2015) explained that women's interest in majoring in engineering in college is greater in engineering departments that offer or frame engineering content that is more societally relevant (e.g., coursework that aims to assist poor communities or reduce inequality). This research leads to our third hypothesis:

Hypothesis 3: Perceptions that a robotics instruction task is good for society will have a stronger positive association with women's interest in the task than with men's interest.

Coupled with the inability of much of STEM education to appeal to women is the association of "brilliance" with the pursuit of STEM disciplines, especially male-dominated ones. For example, fields with the highest expectations of brilliance had the lowest proportions of women, most likely due to stereotypes that women do not have the innate ability to be brilliant in these fields (Leslie et al., 2015; Storage et al., 2016). These last studies form the foundation for the final hypothesis:

Hypothesis 4: Perceptions that a robotics instruction task requires natural talent will have a stronger negative association with women's interest in the task than with men's interest.

METHODS

Data

We collected the data we use to test our hypotheses in August 2017 via a Qualtrics survey administered on Amazon's Mechanical Turk (MTurk). While the use of MTurk in social science experiments remains controversial, MTurk samples have been found to be more diverse than samples recruited on college campuses or via online social media postings (Casler, Bickel & Hackett, 2013; Buhrmester, Kwang, & Gosling, 2011; Paolacci, Chandler & Ipeirotis, 2010). MTurk samples are also just as reliable as in-person samples recruited from college campuses (Casler, Bickel & Hackett, 2013; Buhrmester, Kwang & Gosling, 2011). If they include the right attention checks, MTurk samples are just as reliable as community samples used in classic psychological studies (Rouse, 2015). Although MTurk samples have these benefits, it is important to note that these samples face potential biases because MTurk users tend to be better with technology than the general population.

Participants were eligible for this study if they were from the United States and had an MTurk approval rating of 95 percent or higher. Overall, 154 respondents completed the survey and they were compensated $0.80 for completing this 10-15-minute survey. The responses of twenty-two respondents were dropped from the study because they failed either of the attention checks or the multiple-choice manipulation check question, bringing the final analytic sample total to 132 respondents (63 women, 69 men).

Respondents were asked to read a robotics instruction task scenario (see below) at the beginning of the survey and were asked to answer questions based on that scenario. An experiment was built into the survey by randomly assigning respondents to one of three conditions in which the robotics instruction task scenario was framed as either masculine (programming your robot), feminine (teaching your robot) and gender neutral (training your robot). We used these descriptions because in the United States, "programming" and "teaching" are male-dominated and female-dominated fields or tasks respectively (Kay, Matuszek & Munson, 2015). The survey's robotics instruction task scenario was as follows:

> Throughout this survey, imagine you are given a robot and have to (*program, teach, train*) it to perform various tasks. This (*programming, teaching, training*) will consist of demonstrating tasks for the robot to perform, which does NOT consist of writing any code. You will answer questions about your perceptions of the task of (*programming, teaching, training*) your robot.

To ensure that respondents did not confuse programming, teaching, or training the robot with actual coding, respondents were reminded throughout the survey that the task consisted of demonstrating tasks for the robot to perform, not actually writing the code for the robot to perform these functions.

Attention Checks

We used two attention checks to ensure respondents were paying attention throughout the survey and responding accurately. We dropped respondents from the study for failing either of these attention checks. The first attention check appeared at the beginning of the survey and went as follows:

> In order to collect high quality data, we are also interested in seeing if you take time to read each question carefully. To demonstrate that you have read these instructions, please ignore the question below and click the arrow at the bottom to proceed to the next screen.

What activities do you think robots are able to do as well as humans (check all that apply)?

- ❑ Care for the elderly
- ❑ Compose Music
- ❑ Drive cars
- ❑ Play chess
- ❑ Work a cash register

The second attention check, which appeared at the end of the survey, was as follows:

Realistically, we know some MTurk respondents do not pay close attention to the questions they are answering. This affects the quality of our data. Please select one of the following honestly. Your answer is confidential. It will not affect whether or not you receive payment and will not affect any rating given to you for your work. Did you pay attention and answer honestly?

- o Yes, keep my data
- o No, delete my data

Manipulation Checks

Immediately following the main body of questions at the focus of our analysis, we asked respondents a multiple-choice question about the task they previously carried out in the survey (teaching, training, or programming their robot). This question served as a manipulation check; their answer to this question was used to ensure that respondents detected the manipulation on the survey instrument, but also an assurance that they did not confuse programming, teaching, or training the robot with writing code to program the robot. Respondents were dropped from the study if they failed this manipulation check. This manipulation check was tailored to the framing of the robotics instruction task scenario (teaching, training, or programming) and was worded as follows: In the previous questions, what task were you asked about? And responses included: using code to program robots, playing with robots, teaching robots, or experimenting with robots.

The second manipulation check was measured on a continuum, where respondents were asked to rate the tasks in each study from very feminine to very masculine. Respondents were asked, "Thinking about what you would consider to be stereotypically masculine, feminine, and gender neutral, how would you classify the following tasks?" and could respond: 1=very feminine, 2=somewhat feminine, 3=gender neutral, 4=somewhat masculine, 5=very masculine. The tasks they assessed included: teaching, programming, firefighting, caring for the elderly, taking photos.

We estimated regression models with the experimental group as the independent variable, and each task in the manipulation check (e.g., teaching, programming, firefighting, caring for the elderly, taking photos) as the dependent variable. We found no significant differences in the perceptions of these tasks across the ways we framed the robotics scenario (e.g., as teaching (feminine), training (gender neutral), or programming (masculine) your robot (results available upon request)). We also used t-tests to compare the mean scores of the tasks we would expect the respondents to consider as feminine (teaching and caring for the elderly) and the tasks we would expect to be perceived as masculine (programming and firefighting). In line with our expectations, for the teaching (feminine) versus programming (masculine) contrast, we found that teaching and programming had significantly different mean ratings of femininity to masculinity ($t = -14.039$, $p<.001$). Likewise, firefighting and caring for the elderly had significantly different means for the ratings of femininity versus masculinity ($t = 22.13$, $p<.001$).

Dependent Variables

We use four dichotomous dependent variables to measure interest in each robotics instruction task. Each of the survey items used to measure interest asked the respondent the extent to which they agreed or disagreed with a set of statements and were coded as "1" if they answered Strongly Agree or Agree (high interest) or 0 if they answered Neutral, Disagree, or Strongly Disagree (low interest). We initially tried the coding of "1" for strongly disagree to "5" for strongly agree but used this dichotomous coding because the models violated the proportional odds assumption for ordinal logistic regression. The first item measures the respondent's general interest in the robotics instruction task and is based on the survey item, "In general, it would be fun to [program/teach/train] my robot to perform various tasks." The remaining three dependent variables are based on an item measuring respondents' interest in programming, teaching, or training their robot to perform a masculine, feminine, or gender neutral task. The item was a matrix question beginning with the statement, "It would be fun to program/teach/train my robot to…" and allowed respondents to choose the extent they agreed or disagreed with this statement for the following tasks: fighting small fires (masculine task), performing basic elder care (feminine task), and taking photos with a camera (gender neutral task).

Independent Variables

The independent variables are as follows. First, we measure respondent's gender, a dummy variable (female = 1, male = 0). We then use a series of dummies for the experimental condition of the robotics instruction task scenario presented at the beginning of the survey (programming for masculine, teaching for feminine, and training for gender neutral). We multiply the respondent's gender by the experimental condition to obtain the interaction variables that allow us to assess whether the relationship between the gendered description of the robotics instruction task scenario and interest in each robotics task differs for women and men.

We derive the remaining two independent variables from survey items that indicate the respondent to answer the extent they agree or disagree with a set of statements. First, to see whether a respondent believes that programming, teaching, or training robots is beneficial for society, we asked respondents to answer the extent they agree or disagree with this statement: "There are societal benefits stemming from programming/teaching/training robots to perform tasks for others."

To examine whether a respondent thinks programming, teaching, or training robots requires natural talent, respondents reported the extent they agreed or disagreed to the statement: "Programming/Teaching/Training robots requires natural talent." For these two measures, we code responses as: Strongly Agree = 5, Agree = 4, Neither Agree nor Disagree = 3, Disagree = 2, and Strongly Disagree = 1. We multiply respondent gender by the variables for societal benefits, and natural talent to obtain the interaction variables that allow us to test Hypotheses 3 and 4, that is, whether the relationship between perceptions of the task's societal benefit and natural talent requirements on interest differs for women and men.

Control Variables

We include measures of a respondent's self-rated technology self-efficacy, and locus of control with two scales. These scales were each based on subsets of 5 validated scale items and ranged from strongly agree to strongly disagree (Liou & Kuo, 2014; Valencha & Ostrom, 1974). Items for the technology self-efficacy scale include "Whether content about technology is difficult or easy, I am sure that I can understand it," "If I was being taught about technology, I could understand the concepts very well," "In general, technology topics are easy for me to understand," "I usually do well with technology," and "I can complete difficult work if I try." Interestingly, there were no significant differences in technology self-efficacy between men and women (see Table 1), a result that could occur because the MTurk sample contains skilled users of technology. That is, both female and male MTurk survey takers are tech-savvy.

Items in the locus of control scale include: "When I make plans, I am almost certain that I can make them work," "Chance or luck plays an important role in my life," "In my case, getting what I want has little or nothing to do with luck," "What happens to me is my own doing," and "Getting people to do the right thing depends upon ability; luck has little or nothing to do with it." These scales were highly reliable; technology self-efficacy had a Chronbach's alpha of 0.89, while the locus of control scale alpha score was 0.86.

We also control for respondent race (white = 1, non-white = 0), age (in years, a continuous variable), and education (1 = high school or less, 2 = Some College/Trade Vocational School, 3 = Associates Degree, 4 = Bachelor's Degree or Higher).

Table 1: Means and Standard Deviations for the Full Sample and by Gender

	Total (n=132)	**Women (n=63)**	**Men (n=69)**
Dependent Variables			

Fun – General Robotics Task (Teaching, Training, or Programming a Robot to Perform Tasks)	.76 (.43)	.79 (.41)	.72 (.45)
Fun - Masculine Task (Fighting Fires)	.62 (.48)	.65 (.48)	.61 (.49)
Fun - Feminine Task (Basic Elder Care)	.54 (.50)	.65* (.48)	.43 (.50)
Fun - Gender Neutral Task (Taking a Picture)	.72 (.45)	.76 (.43)	.68 (.47)
Independent Variables			
Gender (Female = 1)	.48 (.50)	-	-
Feminine Condition (Teaching the Robot)	.36 (.48)	-	-
Masculine Condition (Programming the Robot)	.32 (.47)	-	-
Gender Neutral Condition (Training the Robot)	.32 (.47)	-	-
Good for Society	4.11 (.83)	4.11 (.74)	4.10 (.91)
Requires Natural Talent	2.86 (1.03)	2.89 (1.02)	2.83 (1.04)
Controls			
Technology Self-Efficacy Scale	19.64 (3.87)	19.22 (3.55)	20.01 (4.13)
Locus of Control Scale	18.32 (4.12)	17.44* (4.34)	19.15 (3.75)
Education	2.85 (1.11)	-	-
Race (White = 1)	.79 (.41)	-	-
Age	33.39 (9.13)	34.33 (10.16)	32.54 (8.06)

$^*p<.05$

Analytical Technique

Since our dependent variables are binary, we use binomial logistic regression in our analyses. Instead of using the logistic coefficients, we convert them into odds ratios by exponentiating them with base *e* for easier interpretation. Odds ratios less than 1 mean a negative relationship between the variable and outcome, odds ratios greater than 1 mean a positive relationship, and odds ratios equal to 1 mean there is no relationship between the two.

The logistic regression model takes the following form:

$$ln\frac{\pi_i}{1-\pi_i} = X_i\beta_i$$

where $\ln\frac{\pi_i}{1-\pi_i}$ is the logit for high interest, X_i is the design matrix containing a

column of 1's for the constant and columns for each independent variable's observations, and β_i is a vector containing the constant and logistic regression coefficients.

RESULTS

Table 2 presents the results for the first set of binomial logistic regression models, which test whether men will have more interest in the robotics tasks than women (hypothesis 1) and if the framing of a robotics instruction task scenario as "feminine" or "gender neutral" (compared to "masculine") will have a more positive association with women's interest in the robotics tasks than with men's interest (hypothesis 2). There were not issues with multicollinearity because the variance inflation factors (VIFs) were not over 4 for any of the models. Results from the first set of analyses do not support these two hypotheses, but nevertheless yielded interesting patterns of findings. Contrary to our first hypothesis that men would be more likely to be interested in the robotics tasks than women, we found that women were actually twelve times *more* likely to be interested in the general robotics task (OR=12.12; p =.052), about six times as likely to be interested in the masculine robotics task of fighting fires (OR = 6.04; p = .031), and ten times as likely to be interested in the feminine robotics task of elder care (OR = 10.73; p = .005) than were men, net of controls. One possible reason we obtained these unusual results could be that our sample contains men and women who are more skilled users of technology than the overall U.S. population. We discuss other potential reasons for these results in the discussion section.

The description of the robotics instruction task scenario (as either feminine or gender neutral compared to masculine) had almost no relationship with interest in the four different robotics tasks, and when it did, it was in the opposite direction than we hypothesized. To illustrate, the feminine framing (teaching the robot) had a marginally more negative impact on women's likelihood of being highly interested in the general robotics task of teaching the robot (OR = .08; p = .091) than the masculine framing of the task (programming the robot). Similarly, the feminine framing (teaching the robot) of the robotics task scenario had a significantly larger negative impact on women's likelihood of being highly interested in the masculine robotics task (programming the robot) (OR = .11; p = .048). The only significant finding when the task was gender neutral (training the robot) was that this framing was slightly more negatively related to women's likelihood of being highly interested in the feminine robotics task (OR = .14; p = .060) compared to men's.

Table 2: Odds Ratios for the Binomial Logistic Regression Models Predicting Interest in the General, Masculine, and Feminine Robotics Tasks (n = 132; standard errors in parentheses)

	Interest in	Interest in	Interest in

	General Robotics Task	Masculine Task (Fighting Fires)	Feminine Task (Elder Care)
Independent Variables			
Female	12.12^{+}	6.04*	10.73**
	(15.57)	(5.03)	(9.00)
Experimental Condition (Masculine Group of "Programming the Robot" = 0)			
Control Group	.61	1.59	1.06
("Training the Robot")	(.45)	(1.04)	(.66)
Feminine Group	1.35	4.28^{+}	1.21
("Teaching the Robot")	(.15)	3.22	(.78)
Female * Control Group	.15	.18	.14^{+}
	(.22)	(.19)	(.15)
Female * Feminine Group	.08^{+}	.11*	.31
	(.12)	(.12)	(.32)
Controls			
Technology Self-Efficacy	1.33***	1.29***	1.15*
	(.08)	(.09)	(.07)
Locus of Control	1.03	1.00	1.06
	(.07)	(.06)	(.06)
Education	.97	.72^{+}	1.13
	(.20)	(.14)	(.20)
Race	.76	.96	1.26
	(.48)	(.50)	(.61)
Age	1.00	.99	.97
	(.03)	(.02)	(.02)
Pseudo R-Squared	.20	.15	.14

$^{+}$*p<.10; *p<.05; **p<.01; ***p<.001*

Table 3 presents the results testing hypotheses 3 (perceptions that robotics tasks are good for society will have a stronger, positive association with women's interest in these task than men's) and hypothesis 4 (perceptions that robotics tasks require natural talent will have a stronger, negative association with women's interest in these tasks than men's). We find limited support for hypothesis 3; considering the feminine robotics task as good for society was only marginally more positively related to women's interest in the feminine robotics task of elder care than men's (OR = 6.04, p = .096). However, those who considered robotics tasks as good for society – regardless of gender – were about four times as likely to be interested in the general robotics task (OR = 4.38, p = .065) and six times as likely to be interested in the masculine robotics task of fighting fires (OR = 6.31, p = .038).

None of the gender and natural talent evaluation interactions were statistically significant, so we do not include these in models (results available upon request). These non-significant gender interactions imply that the relationship between perceptions that a task requires natural talent on interest in the task is similar for women and men, so we find no support for hypothesis 4.

Table 3: Odds Ratios for the Binomial Logistic Regression Models Predicting Interest in the General, Masculine, Feminine, and Gender Neutral Robotics Tasks (n = 132; standard errors in parentheses)

	Interest in General Robotics Task	Interest in Masculine Task (Fighting Fires)	Interest in Feminine Task (Elder Care)	Interest in Gender Neutral Task (Photography)
Independent Variables				
Female	1.28	2.34	.76	1.68
	(1.24)	(2.53)	(.71)	(1.80)
Good for Society	4.38^{+}	6.31*	.84	5.91*
	(3.51)	(5.62)	(.62)	(5.33)
Female * Good for	1.61	.59	6.04^{+}	.98
Society	(1.86)	(.70)	(6.53)	(1.21)
Controls				
Technology Self-Efficacy	1.24***	1.20**	1.13*	1.30***
	(.08)	(.07)	(.07)	(.10)
Locus of Control	1.00	.98	1.06	.86*
	(.07)	(.06)	(.06)	(.07)
Education	1.13	.80	1.21	1.09
	(.52)	(.15)	(.22)	(.23)
Race	.85	.97	1.19	.12*
	(.52)	(.49)	(.56)	(.10)
Age	.98	.98	.96	1.05^{+}
	(.03)	(.02)	(.02)	(.03)
Pseudo R-Squared	.21	.16	.13	.27

$^{+}$*p<.10; *p<.05; **p<.01; ***p<.001*

DISCUSSION

Overall, the findings from our analyses contradict our hypotheses. First, contrary to hypothesis 1, we found that women were significantly *more* likely to have interest in most of the robotics tasks compared to men. What is more, we found that a "feminine" framing of the robotics task scenario (e.g., teaching a robot) had a significantly higher negative impact on women's likelihood of being highly interested in the masculine robotics task. We observed limited support for the third hypothesis, which predicted that perceiving robotics tasks as good for society would have a stronger impact on women's interests in the robotics tasks than on men's interest. Finally, results do not support our final hypothesis, which predicted that perceiving robotics tasks as requiring natural talent would have a stronger impact on women's interest in the robotics tasks than men.

While our results did not support our hypotheses, our study raises interesting questions related to research on gender and STEM interest more broadly. One possible explanation for our unusual finding that women are more interested in the robotics instruction task than men, is that women really are more interested in working with the robots given the way we framed our study and measured interest in technology. As shown in the regression models, those who considered several of the robotics learning tasks to be beneficial to society – regardless of gender or the framing of the robotics learning task as feminine, masculine, or gender neutral – were significantly more likely to be interested in these tasks. Although we did not find evidence that thinking a robotics task is beneficial to society had a stronger impact on women's interests than men's interests in the robotics tasks, several studies have found that women are interested in doing science that impacts society (Smith-Doerr et al., 2016; Espinosa, 2011). Said differently, the women in our sample may want to engage in technology that helps society and the tasks we ask about align with this interest in helping society.

It is also possible that the robotics instruction task scenario itself – regardless of its framing as teaching, training, or programming – is interpreted as more interesting for women than men because robotics learning itself is very similar to subfields of computer science where women have achieved gender parity in representation, such as human-computer interaction (Dray et al., 2013). One study found that about 30 percent of bachelor's degrees in interdisciplinary computer science fields (which include the subfield of human-computer interaction) were earned by women in 2013, but only about 15 percent of computer science bachelor's degrees were earned by women this same year (Zweben & Bizot, 2016). Women's higher interest in this robotics task could also stem from the technical/social divide that is present in engineering (Cech, 2013; Cech, 2014); since the robotics task is more practical and involves more interaction than a more theoretical or technical task, this could explain why women in our sample expressed greater interest than men.

Another reason we reach a different conclusion than earlier studies with regard to women's interest in robotics tasks compared to men may reflect the detail of our design. That is, our study measured interest in a different context than prior studies looking at gender and interest in computer science. To illustrate, we asked respondents about specific interest in hypothetically teaching, training, or programming robots to perform various tasks. In other words, we asked respondents to consider actually engaging in computer science related tasks whereas other studies have asked about broader interest in computer science or interest in taking computer science courses (e.g., Master, Cheryan & Meltzoff 2016). It is possible that specifying the task one would actually perform with a robot, rather than asking for general interest in computer science, led to the differences in findings between our study and prior research (e.g., Cheryan et al., 2009).

Ultimately, our findings pose a methodological warning for future researchers interested in studying the gender divide in computer science. Researchers should pay close attention to subfields – whether of computer science or other STEM fields – when assessing women's interest or engagement. Others have pointed out this criticism, noting that women's representation across STEM fields is not equal; for example, women make up about 40 percent of mathematics bachelor degree holders, but about 15 percent of computer science bachelor's degrees (National Science Foundation, 2016). Researchers should also consider measuring interest in STEM by inquiring about interests in learning or performing the specific tasks they would actually perform in an academic or industry work setting, not simply by asking about a general interest in STEM fields (e.g., "How interested are you in majoring in computer science?"). It is possible that how you measure interest in STEM could impact gender similarities or differences in interest.

Alternatively, the lack of support for our hypotheses could stem from our study design. First, it is possible that "teaching" (the term we used to signal a feminine task) is not overwhelmingly liked by men *or* women. At least in the United States, the teaching profession is associated with low value and little pay. Several studies in the social science literature back up the notion that women's work is devalued, especially professions like teaching (Rubery, 2017; Rutherford, 2011). Lastly, considering that the analytic sample is from MTurk, which requires the use of the computer (albeit very basic level computer work), it is possible that our observation that women are more interested in the robotics tasks than men because the women taking part in the survey experiment feel more comfortable with technology than other women in the U.S. population.

LIMITATIONS

This study has several limitations. First, we only asked about perceptions of different robotics tasks and did not actually have respondents perform the tasks. It is possible that respondents would behave differently towards the robotics tasks (in terms of interest) had they actually performed them. Similarly, the robotics scenario and tasks were not necessarily relatable to everyday life or coursework. We may have observed more differences in perceptions of these had the respondents actually completed robotics tasks on the job or in more realistic settings. The robotics scenario and task were also removed from everyday structural contexts, such as work organizations or STEM classrooms. Since computer science and technology related fields are male dominated, this context is important to consider.

These results also may not be generalizable to the entire U.S. population. Although we used random assignment by experimental condition, we did not randomly *sample* respondents from the U.S. population, so we cannot say these results are representative of the U.S. population at large. The MTurk sample itself may misrepresent the U.S. population because MTurk users are a self-selected group who used a computer to complete the survey, and hence may be more skilled with technology than the average American. This characteristic of the sample could be one reason that women actually found the robotics tasks to be more interesting than men. We also had a small sample. A larger, more representative sample could have led to different results.

FUTURE RESEARCH AND IMPLICATIONS

This study sets up several avenues for future research and replication studies. First, to address the possible concern that the term "teaching" has a negative connotation with women and men, future research should experiment by using different ways to frame a robotics task scenario as feminine (e.g., "guiding" the robot). Second, future studies could use experimental methods to have participants actually do a real-life robotics learning task (framed as masculine, feminine, or gender neutral) and see how this impacts both performance and interest in the task. It is possible that examining a real-life task would yield different results than those found in our study. A real-life task would also have the advantage of providing additional context to the robotics task that we could not provide in the study. Third, future research could explore whether setting the robotics learning task in a male-dominated context (e.g., computer science course) versus a female-dominated context (e.g., hospital) would matter. That is, whether gender segregation in combination with gendered framing impacts perceptions of interacting with robots.

We conclude by describing several practical applications for this research. By learning about women's and men's interest in various robotics tasks, these data and data from future studies could help people learn to interact with robots more effectively and become better at training robots to perform tasks. Our findings have implications for STEM education and work organizations. Even though our findings did not align entirely with our hypotheses, we know in certain scenarios that women can have more interest in technological tasks than men and that those who consider robotics tasks as beneficial to society – regardless of gender – also are more likely to have high interest in technological tasks. Figuring out the mechanisms behind these findings and implementing programs and practices around them will help diversify STEM education and the STEM workforce.

ACKNOWLEDGEMENTS
The authors thank Cynthia Matuszek for feedback on the Mechanical Turk survey.

REFERENCES

Appianing, J., & Van Eck, R. N. (2015). Gender differences in college students' perceptions of technology-related jobs in computer science. *International Journal of Gender, Science and Technology*, *7*(1), 28-56.

Beyer, S. (2014). Why are women underrepresented in Computer Science? Gender differences in stereotypes, self-efficacy, values, and interests and predictors of future CS course-taking and grades. Computer Science Education, 24(23), 153-192.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data?. *Perspectives on Psychological Science*, *6*(1), 3-5.

Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior*, *29*(6), 2156-2160.

Cech, E. A. (2013). Ideological wage inequalities? The technical/social dualism and the gender wage gap in engineering. *Social Forces*, *91*(4), 1147-1182.

Cech, E. A. (2014). Culture of disengagement in engineering education?. *Science, Technology, & Human Values*, *39*(1), 42-72.

Chernova, S., & Thomaz, A. L. (2014). Robot learning from human teachers. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, *8*(3), 1-121.

Cheryan, S., Plaut, V. C., Davies, P. G., & Steele, C. M. (2009). Ambient belonging: how stereotypical cues impact gender participation in computer science. *Journal of Personality and Social Psychology*, *97*(6), 1045-1060.

Davison, C., & Argyriou, E. (2016). Gender Preferences in Technology Adoption: An Empirical Investigation of Technology Trends in Higher Education. *International Journal of Gender, Science and Technology*, *8*(3), 405-419.

Dray, S. M., Peters, A. N., Brock, A. M., Peer, A., Druin, A., Gitau, S., Kumar, J., & Murray, D. (2013). Leveraging the progress of women in the HCI field to address the diversity chasm. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems* (pp. 2399-2406). ACM.

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: a meta-analysis. *Psychological bulletin*, *136*(1), 103-127.

Espinosa, Lorelle. (2011). "Pipelines and Pathways: Women of Color in Undergraduate STEM Majors and the College Experiences that Contribute to Persistence." *Harvard Educational Review* 81(2): 209-241.

Eyssel, F., & Hegel, F. (2012). (S)he's got the look: gender stereotyping of robots. *Journal of Applied Social Psychology*, *42*(9), 2213-2230.

Galpin, V. (2002). Women in computing around the world. ACM SIGCSE Bulletin, 34(2), 94-100.

Kay, M., Matuszek, C., & Munson, S. A. (2015). Unequal representation and gender stereotypes in image search results for occupations. In *Proceedings of the 33rd*

Annual ACM Conference on Human Factors in Computing Systems (pp. 3819-3828). ACM.

Knox, W. B., & Stone, P. (2008). Tamer: Training an agent manually via evaluative reinforcement. In *Development and Learning, 2008. ICDL 2008. 7th IEEE International Conference on* (pp. 292-297). IEEE.

Kuchenbrandt, D., Häring, M., Eichberg, J., Eyssel, F., & 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.

Kuhlmann, G., Stone, P., Mooney, R., & Shavlik, J. (2004, July). Guiding a reinforcement learner with natural language advice: Initial results in RoboCup soccer. In *The AAAI-2004 workshop on supervisory control of learning and adaptive systems*.

Lagesen, V. A. (2008). A cyberfeminist utopia? Perceptions of gender and computer science among Malaysian women computer science students and faculty. *Science, Technology, & Human Values*, *33*(1), 5-27.

Leslie, S. J., Cimpian, A., Meyer, M., & Freeland, E. (2015). Expectations of brilliance underlie gender distributions across academic disciplines. *Science*, *347*(6219), 262-265.

Liou, P. Y., & Kuo, P. J. (2014). Validation of an instrument to measure students' motivation and self-regulation towards technology learning. *Research in Science & Technological Education*, *32*(2), 79-96.

Loftin, R., Peng, B., MacGlashan, J., Littman, M. L., Taylor, M. E., Huang, J., & Roberts, D. L. (2016). Learning behaviors via human-delivered discrete feedback: modeling implicit feedback strategies to speed up learning. *Autonomous Agents and Multi-Agent Systems*, *30*(1), 30-59.

Master, A., Cheryan, S., & Meltzoff, A. N. (2016). Computing whether she belongs: Stereotypes undermine girls' interest and sense of belonging in computer science. *Journal of Educational Psychology*, *108*(3), 424.

Miller, D. I., Eagly, A. H., & Linn, M. C. (2015). Women's representation in science predicts national gender-science stereotypes: Evidence from 66 nations. *Journal of Educational Psychology*, *107*(3), 631.

Mutlu, B., & Forlizzi, J. (2008). Robots in organizations: the role of workflow, social, and environmental factors in human-robot interaction. In *Human-Robot Interaction (HRI), 2008 3rd ACM/IEEE International Conference on* (pp. 287-294). IEEE.

Nass, C., Moon, Y., & Green, N. (1997). Are machines gender neutral? Gender-stereotypic responses to computers with voices. *Journal of Applied Social Psychology*, *27*(10), 864- 876.

National Science Foundation. (2016). Bachelor's degrees awarded, by sex and field: 2004–2014. Retrived from https://www.nsf.gov/statistics/2017/nsf17310/static/data/tab5-2.pdf

Nilsson, L. 2015. How to attract female engineers. *The New York Times.* Retrieved August 24, 2018 from:

https://www.nytimes.com/2015/04/27/opinion/how-to-attract-female-engineers.html

Othman, M., & Latih, R. (2006). Women in computer science: no shortage here!. *Communications of the ACM*, *49*(3), 111-114.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk.

Ridgeway, C. (2011). Framed by Gender: How Gender Inequality Persists in the Modern World. Oxford University Press.

Riegle-Crumb, C., King, B., Grodsky, E., & Muller, C. (2012). The more things change, the more they stay the same? Prior achievement fails to explain gender inequality in entry into STEM college majors over time. *American Educational Research Journal*, *49*(6), 1048-1073.

Rouse, S. V. (2015). A reliability analysis of Mechanical Turk data. *Computers in Human Behavior*, *43*, 304-307.

Rubery, J. (2017). Why is Women's Work Low-Paid? Establishing a framework for understanding the causes of low pay among professions traditionally dominated by women. Working Paper.

Rutherford, S. (2011). *Women's work, men's cultures: overcoming resistance and changing organizational cultures*. Palgrave Macmillan.

Smith-Doerr, Laurel, Vardi, I., & Croissant, J. (2016). "Doing Gender and Responsibility: Scientists and Engineers Talk About Their Work." *Journal of Women and Minorities in Science and Engineering 22*(1): 49-68.

Storage, D., Horne, Z., Cimpian, A., & Leslie, S. J. (2016). The frequency of "brilliant" and "genius" in teaching evaluations predicts the representation of women and African Americans across fields. *PloS one*, *11*(3), 1-17, e0150194.

Thomaz, A. L., & Breazeal, C. (2006, July). Reinforcement learning with human teachers: Evidence of feedback and guidance with implications for learning performance. In *Aaai* (Vol. 6, pp. 1000-1005).

Valencha, G.K. & Ostrom, T.M. (1974). An abbreviated measure of internal-external locus of control. *Journal of Personality Assessment*, *38*(4), 369–376.

Witt, N., & Hofmeister, H. (2015). "I just have to have it" or "It's enough for me"? Gendered tendencies in attitudes towards and usage of mobile communication technology. *International Journal of Gender, Science and Technology*, *7*(1), 4-27.

Wynn, A. T., & Correll, S. (2018). "Puncturing the Pipeline: Do Technology Companies Alienate Women in Recruiting Sessions?" *Social Studies of Science*. *48*(1), 149-164.

Zweben, S. H., & Bizot, E. B. (2016). Representation of women in postsecondary computing: Disciplinary, institutional, and individual characteristics. *Computing in Science & Engineering*, *18*(2), 40-56.