Do Robots Need to Be Stereotyped? Technical Characteristics as a Moderator of Gender Stereotyping

Florian Dufour * and Céline Ehrwein Nihan

Interdisciplinary Institute of Business Development, School of Engineering and Management Vaud (HEIG-VD), University of Applied Sciences of Western Switzerland (HES-SO), Avenue des Sports, 20, 1401 Yverdon-les-Bains, Switzerland; celine.ehrwein@heig-vd.ch
* Correspondence: florian.dufour@heig-vd.ch; Tel.: +41-024-557-6170

Abstract: As suggested by previous results, whether, when designing robots, we should make use of social stereotypes and thus perpetuate them is question of present concern. The aim of this study was the identification of the specific conditions under which people's judgments of robots were no longer guided by stereotypes. The study participants were 121 individuals between 18 and 69 years of age. We used an experimental design and manipulated the gender and strength of robots, and we measured the perception of how a robot could be used in automotive mechanics for light and heavy tasks. Results show that the technical characteristics of robots helped to anchor people's judgments on robots' intrinsic characteristics rather than on stereotypical indicators. Thus, stereotype perpetuation does not seem to be the sole option when designing robots.

1. Introduction

About fifty years ago, Isaac Asimov [1] imagined what a World's Fair in 2014 would look like. Among other things, Asimov described a "robot housemaid [...] slow-moving but capable of general picking-up, arranging, cleaning, etc." The attribution of a human gender stereotype to a (gender-free) machine seems odd, so why could one imagine this "female" robot with great ease?

Media equation theory [2] helps in answering this question. This theory assumes that processes guiding social judgments and interactions can be generalized to encompass human–media (including human-computer and human-robot) judgments and interactions. For instance, using the computers are social actors experimental paradigm, [3] showed how common social responses to computers are. Concerning robots specifically, several empirical studies have shown that the perception of robots can be driven by social categorization processes and stereotypes. For instance, a "male" robot has been perceived as more agentic [4], more competent to achieve male stereotyped tasks [4], or more suitable for use as a security guard [5] than a "female" robot. To say the least, it seems that judgments of robots are not immune to stereotyping, even if the picture is not that simple, given the results of a recent study [6] showing that gender stereotyping of robots has little, if any, effect on such judgments. Despite the need to have a better understanding of this asymmetry of results concerning gender stereotypes, it seems that, at least under some circumstances, human–robot interactions or people's judgments of robots involves social categorization processes and stereotyping. Hence, it could be tempting to build social robots [7] embedding social characteristics stereotypically congruent with their intended use (e.g., a "male" security guard) to enhance their acceptance and economic value; however, this would come at the price of stereotype perpetuation. In this study, we argue that socio-economic and moral principles do not necessarily need to be in conflict.

In previous studies, the role of technical characteristics associated with stereotypical characteristics was not studied in depth. We assume that the technical characteristics of a robot congruent with success in performing a task should moderate the effect of gender stereotypes. In other words, even if gender roles are so embedded in the human psyche (e.g., [8]) that they will guide the evaluation of robots in accordance with stereotypes, people should also rely on the objective skills or attributes of robots, when available, to form more accurate judgments [9]. More specifically, in the context of the reflection-construction model of Lee Jussim [10], when no objective characteristics (congruent with a success on a task) of robots are known, social beliefs about gender will have a strong influence on the perceiver's judgments; however, when target's attributes are known (and congruent with a success on a task), social beliefs will lose their explanatory strength and people will use the known objective attribute when judging robots.

Consequently, we formulated one general interaction hypothesis and two specific hypotheses bringing our predictions in line with the reflection–construction model.

General hypothesis: The effect of gender stereotype on judgments of a robot is moderated by the robot's characteristics.

Hypothesis 1. When the technical characteristic of a robot is unknown or insufficient to achieve a task, gender stereotypes guide judgments (see Table 1).

Table 1. Predictions drawn from the hypotheses.

Referring Theoritical Hypothesis	Robot's Characteristic	Type of Tasks	Robot's Characteristic Sufficient to Achieve the Task	Expected Result	Corresponding Effects (see Section 3)
Hypothesis 1	No information	Heavy tasks	Unknown	Stereotyping (Male > Female)	Effect 1, 2
Hypothesis 1	No information	Light tasks	Unknown	Stereotyping (Male > Female)	Effect 1, 2
Hypothesis 1	Maximum lift 15 kg	Heavy tasks	No	Stereotyping (Male > Female)	Effect 6, 7
Hypothesis 2	Maximum lift 15 kg	Light tasks	Yes	No stereotyping (Male = Female)	Effect 5, 6
Hypothesis 2	Maximum lift 150 kg	Heavy tasks	Yes	No stereotyping (Male = Female)	Effect 3, 4
Hypothesis 2	Maximum lift 150 kg	Light tasks	Yes	No stereotyping (Male = Female)	Effect 3, 4

Hypothesis 2. When the technical characteristic of a robot is known and is sufficient to achieve a task, the observer relies on technical characteristics rather than on a gender indicator (see Table 1).

Current Study

To test our hypotheses, we chose to study an unmitigated male stereotyped activity in the cultural context of this study, i.e., mechanics. We used the ability to lift weights (maximum lift of 15 kg vs. 150 kg) as the known technical characteristic of the robot so that we could have conditions where people could explicitly consider objective skills when judging robots. Moreover, we used two different types of mechanical tasks—those requiring great strength (heavy tasks, for instance, changing a motor) and those requiring little strength (light tasks, for instance, changing spark plugs)—to create conditions where objective characteristics have different implications on success or failure on a task. In doing so, we were able to evaluate the specific conditions under which objective skills are truly taken into account—in other words, where people do or do not judge stereotypically. More specifically, the predictions listed in Table 1 are drawn from the above hypotheses.

2. Materials and Methods

2.1. Participants

Participants were 121 adult visitors (42 women, 78 men, 1 unknown, *mean age* = 31.43, *SD* = 13.48, ranging from 18 to 69) of the Open Doors event at HEIG-VD who volunteered to participate and who answered correctly to (memory) check questions (see below). None of the participants declared they were aware of the variables we were manipulating.

2.2. Procedure

The Open Doors event is a single afternoon event taking place once a year at HEIG-VD. Experimenters held a stand and asked visitors if they wanted to fill in a computerized questionnaire (questionnaires were divided into two unrelated parts, with the first part dedicated to ubiquitous computing in the workplace and the second part dedicated to this study). We used Limesurvey, an online survey open source software, to implement the study. Participants who volunteered received candies to thank them for their participation.

Before filling in the questionnaire, participants completed a consent form. They were informed that their responses were anonymous and that they could withdraw at any time, and they were also asked if they agreed or not to the researchers using their data in scientific publications or in teaching.

On the first webpage of the questionnaire, we manipulated our independent variables and measured their impact on the dependent variable. The second webpage contained manipulation (memory) check questions. We used a 2 (robot's gender: male vs. female, between-subject) × 3 (maximum lift: 15 kg vs. 150 kg vs. no information, between-subject) × 2 (type of tasks: heavy vs. light, within subject) mixed design. A php script randomly directed participants to experimental conditions, a technique which was used so that none of the participants or experimenters were aware of the experimental conditions (double blind process).

To manipulate the robots' gender, we used either a "male version" of FLobi (see [11], the author of which owns the copyright for the pictures) named "Christian", or a "female version" named "Christine". Below the picture located on the top of the first webpage, participants read the text containing our manipulation of the robot's strength (i.e., the robot's maximum lift, in italics below): "[Christine/Christian] is currently under development and will incorporate numerous features. For instance: basic computations and automated task execution, vocal recognition [*lifts weights up to 15 kg* vs. *lifts weights up to 150 kg* vs. *no information*]". Then, participants had to indicate to which extent the robot could be used in automotive mechanics for (a) "heavy tasks (for example: motor, wheel changing)" and (b) "light tasks (for example: spark plugs, mirrors changing)" (within-subjects independent variable) on a scale ranging from "1 = strongly disagree" to "7 = totally agree". Finally, on the second webpage, participants were asked to answer some manipulation (memory) check questions (asking for the gender and the maximum lift for the robot they saw). Check questions were used to ensure that participants carried out the experiment using the instructions they received during the experiment (to eliminate some noise from the data, that is, for instance, participants who thought the robots they had seen was a male when it was actually Christine).

3. Results

3.1. Analysis

We used a series of multiple linear regressions to analyze our 2 (robot's gender: male vs. female, between-subject) × 3 (maximum lift: 15 kg vs. 150 kg vs. no information, between subject) × 2 (type of tasks: heavy vs. light, within subject) design. More specifically, we directly used contrast and dummy coding to analyze our data because (a) our hypotheses were specifics (e.g., simple interaction effects) and (b) main and global interactions effects were of less interest.

More specifically, we tested for seven effects to address our hypotheses.

Hypothesis 1: stereotyping.

- Effect 1: effect of gender where there is no information on strength (we expect this effect to be significant to replicate stereotype effect for both heavy and light tasks).
- Effect 2: interaction effect between gender and the type of tasks where there is no information on strength (we expect to find no significant interaction effect revealing that Effect 1 is identical for both types of tasks).
- Effect 7: effect of gender where strength was explicitly insufficient to achieve the tasks, i.e., a robot with the ability to lift a maximum of 15 kg had to lift a weight far an excess of this (we expected this effect to be significant).

Hypothesis 2: no stereotyping.

- Effect 3: effect of gender where strength was explicitly sufficient to achieve both tasks (we expected this effect to be non-significant).
- Effect 4: interaction effect between gender and the type of tasks where strength was explicitly sufficient, i.e., a robot with the ability to lift 150 kg for heavy and light tasks (we expected to find no significant interaction revealing that Effect 3 is identical for both types of tasks).
- Effect 5: effect of gender in the only other condition where strength was explicitly sufficient to achieve the tasks, i.e., a robot with the ability to lift a maximum of 15 kg had to lift a light weight (we expected this effect to be non-significant).
- Effect 6: interaction effect between gender and the type of tasks in the condition of the maximum lift of 15 kg, because strength, in one case, is explicitly sufficient to achieve the tasks (see Effect 5) and, in the other case, explicitly insufficient to achieve the tasks (see Effect 7) (we expected this effect to be significant as an indication of the moderation of the stereotype effect by the technical characteristics when known and sufficient to achieve the tasks).

To test these effects, we followed the recommendations of Judd, McClelland, and Ryan [12] for analyzing mixed designs. We first computed standardized differences variables ("W_{ki}" see Equation (1) and Table 2 below) to handle non-independence of observations given our within-subjects independent variable, and we then used specific codes to test for each effect of interest (see Equation (2) and Table 2). To eliminate this non-independence problem, given the linear regression framework of our analysis, we computed, for instance, a single composite score for each participant consisting, broadly speaking, of a difference score between the evaluation of the robot for the heavy tasks and the evaluation of the robot for the light tasks (see W_{2i} below). In other words, the dependence of observations was no longer a problem since the composite variable represents the difference within each participant. For instance, if using contrast codes of +1 and −1 respectively for heavy and light tasks in Equation (1), a resulting score of 0 indicates no difference; a positive score indicates a higher score for heavy tasks than light tasks, and a negative score indicates a lower score for heavy tasks than light tasks. Testing the main effect of this variable can now be easily achieved using a regression model (for instance, in an intercept-only model with this composite variable as the response variable and no explanatory variable, the result and p-value are strictly identical to: (a) a one sample *t*-test against a test value of 0; (b) a dependent *t*-test comparing evaluations between heavy and light tasks; or (c) a repeated measure ANOVA with the type of tasks as within-subject independent variable). It is important to note that we used a somewhat more complicated formula instead, for instance, of a difference score to "standardize" scores in order to keep them in the same metric as Y scores rather than W scores. The numerator part is used to calculate the differences between variables and the denominator to standardize scores.

Table 2. Codes used for $Code_z$ explanatory variables in Equation (2) given levels of gender of the robot and the lift independent variables.

Independent Variables		$Code_z$		
Gender of robot	Lift	Code1	Code2	Code3
Female	No information	0	0	1
Female	15 kg	1	0	0
Female	150 kg	0	1	0
Male	No information	0	0	1
Male	15 kg	1	0	0
Male	150 kg	0	1	0
		To test effect 1, 2, 3, 4	To test effect 1, 2, 5, 6, 7	To test effect 3, 4, 5, 6, 7

We used the following equation to calculate standardized differences variables:

$$W_{ki} = \frac{\sum_h \delta_h Y_{hi}}{\sqrt{\sum_h \delta_h^2}} \quad (1)$$

We applied Equation (1) and calculated four dependent variables as follows to test for specific effects (see hypothesis testing subsection below). Because we are using a model comparison approach and a different coding scheme to test for our specific effects, we have to compute four different dependent variables to be used in regression models:

(a)　W_{1i} to test Effects 1 and 3 with $\delta_{1=heavy} = 1$ and $\delta_{2=light} = 1$ with this coding scheme; because δ_1 and δ_2 take the same value, the effect of the type of tasks is no longer of interest (no variation due to the levels of the types of tasks); in other words, the effects of other variables in the model (gender and maximum lift) are tested for both tasks.

(b)　W_{2i} to test Effects 2, 4, and 6 with $\delta_{1=heavy} = 1$ and $\delta_{2=light} = -1$ with this coding scheme; because δ_1 and δ_2 are of opposite sign, the scores of W2 are difference scores, so we can take into account the effect of the type of tasks. Using this variable in the model allows one to estimate the effect of the type of tasks and its interactions with other variables.

(c)　W_{3i} to test Effect 5 with $\delta_{1=heavy} = 1$ and $\delta_{2=light} = 0$ with this coding scheme; because δ_1 takes value of 0, the effect of other variables will only be tested on the light condition, which is similar to analyzing the light condition alone without taking into account the heavy condition (i.e., simple effect).

(d)　W_{4i} to test Effect 7 with $\delta_{1=heavy} = 0$ and $\delta_{2=light} = 1$ with this coding scheme; because δ_2 takes value of 0, the effect of the other variables will only be tested on the heavy condition; which is similar to analyzing the heavy condition alone without taking into account the light condition (i.e., simple effect).
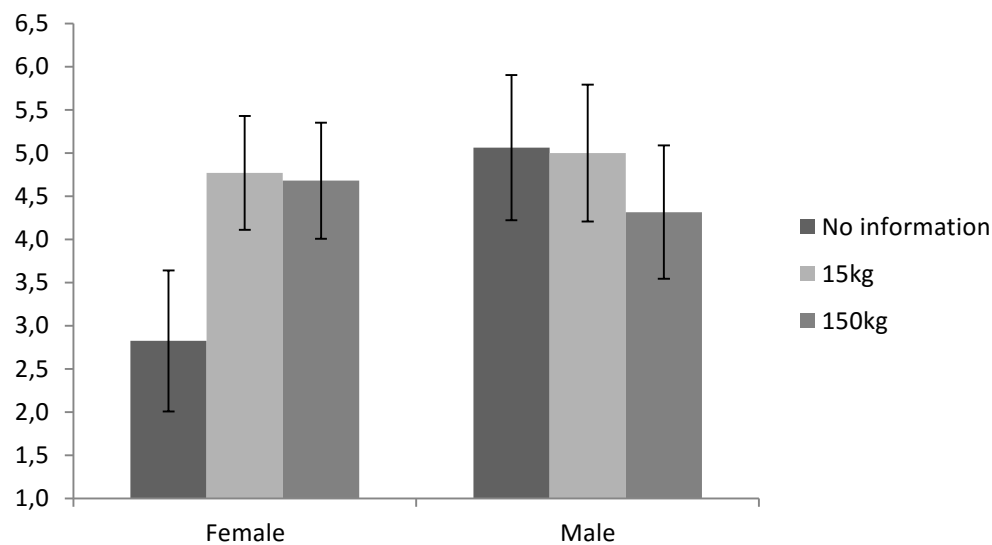
Linear regression model:

$$W_{ki} = \beta_0 + \beta_1 Gender_i + \beta_2 Code_{zi} + \beta_3 Code_{zi} + \beta_4 Gender_i * Code_{zi} + \beta_5 Gender_i * Code_{zi} + \varepsilon_{ki} \quad (2)$$

To predict our dependent variables, we used a contrast code for the gender of the robot—an independent variable (male = 1 vs. female = −1)—and codes listed in Table 2 for the two other explanatory variables in Equation (2):

3.2. Hypotheses Testing

Results are shown in Figures 1 and 2.

Figure 1. Mean of agreement with statements concerning the use of robots in automotive mechanics for light tasks as a function of the robot's gender, and the robot's strength (maximum lift). Error bars represent 95% confidence intervals.
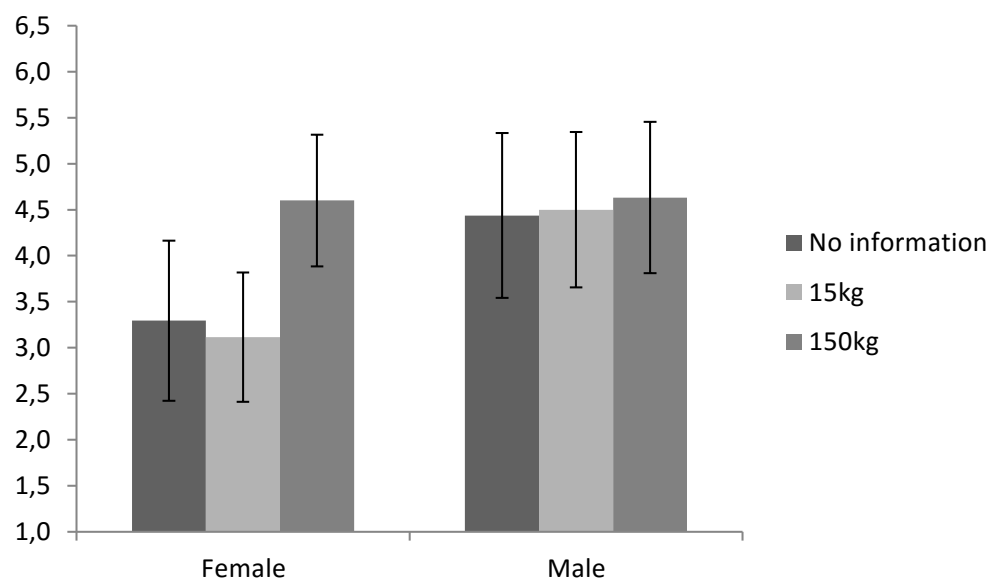
Figure 2. Mean of agreement with statements concerning the use of robots in automotive mechanics for heavy tasks as a function of the robot's gender, and the robot's strength (maximum lift). Error bars represent 95% confidence intervals.

In this section, we report regression coefficients in the metric of the dependent variable in dividing coefficients by the denominator of W_k. We analyzed these data with a robust linear model and found very similar results (the authors will be pleased to provide results of those analyses to the interested reader). Given this similarity and because sphericity was assumed, we were reasonably confident that the results of multiple linear regressions we reported in this results section were not affected by assumption violations and potential outliers.

When no information on strength was given, we replicated the stereotypes' effect (Hypothesis 1) with a significant effect of gender, $b = 0.85$, $t(115) = 3.52$, $p < 0.001$, and *Cohen's d* = 0.67 (Effect 1), in the expected direction. Moreover, we found no significant interaction between gender and the type of tasks, $b = -0.28$, $t(115) = -1.45$, $p = 0.15$, $d = 0.29$ (Effect 2). The participants found the male

robot as more likely to be used in automotive mechanics than the female robot both for heavy tasks (M^{female} = 3.29, SD^{female} = 1.90, 95% CI = [2.42, 4.16] vs. M^{male} = 4.44, SD^{male} = 2.00, 95% CI = [3.54, 5.33]) and light tasks (M^{female} = 2.82, SD^{female} = 1.20, 95% CI = [2.01, 3.64] vs. M^{male} = 5.06, SD^{male} = 1.81, 95% CI = [4.22, 5.90]).

To test Hypothesis 2, we first tested for the gender effect in conditions where strength was explicitly sufficient to achieve both tasks (i.e., 150 kg of strength). We found no effect of gender, $b = -0.08$, $t(115) = -0.40$, $p = 0.69$, $d = 0.07$ (Effect 3) and no significant interaction between gender and the type of tasks, $b = 0.10$, $t(115) = 0.60$, $p = 0.55$, $d = 0.11$ (Effect 4). The participants did not find the male robot more likely to be used in automotive mechanics than the female robot either for heavy (M^{female} = 4.60, SD^{female} = 1.63, 95% CI = [3.88, 5.32] vs. M^{male} = 4.63, SD^{male} = 1.92, 95% CI = [3.81, 5.45]) or light tasks (M^{female} = 4.68, SD^{female} = 1.49, 95% CI = [4.01, 5.35] vs. M^{male} = 4.32, SD^{male} = 2.11, 95% CI = [3.54, 5.09]).

Then, we tested for the effect of gender in the only other conditions where strength was sufficient to achieve the tasks (i.e., 15 kg for light tasks). Interestingly, we did not find any effect of gender (M^{female} = 4.77, SD^{female} = 1.68, 95% CI = [4.11, 5.43] vs. M^{male} = 5.00, SD^{male} = 1.81, 95% CI = [4.21, 5.79]), $b = -0.12$, $t(115) = -0.44$, $p = 0.66$, $d = 0.09$ (Effect 5).

To summarize, when strength was sufficient to achieve the tasks, the effect of stereotypes did not appear. Moreover, the interaction between gender and the type of tasks in the 15-kg condition was tendentiously significant with a small to medium size, $b = 0.29$, $t(115) = 1.74$, $p = 0.08$, $d = 0.35$ (Effect 6), but clearly in the expected direction. Indeed, contrary to the condition of 15 kg for light tasks, in the case where strength was explicitly insufficient to achieve the tasks (i.e., 15 kg for heavy tasks), a significant effect of gender, in the expected direction, was found (M^{female} = 3.12, SD^{female} = 1.66, 95% CI = [2.41, 3.82] vs. M^{male} = 4.50, SD^{male} = 1.89, 95% CI = [3.65, 5.35]), $b = 0.69$, $t(115) = 2.49$, $p = 0.01$, $d = 0.41$ (Effect 7). This latter result was consistent with Hypothesis 1.

A visual inspection of means and their confidence intervals revealed an unintended result. In the male robot/heavy tasks condition (see Figure 2), we did not find a significant difference between the 15-kg condition, where strength was insufficient to achieve the tasks ($M = 4.50$, $SD = 1.89$), and the 150-kg condition, where strength was sufficient to achieve the tasks ($M = 4.63$, $SD = 1.92$).

4. Discussion

In recent studies, robots have been found to be, like their creators, subject to social stereotypes (e.g., [4]). Does the acceptance and hence economic value of robots need to be grounded in stereotypes? We have given a negative answer to this question in showing that, when the technical characteristics of robots are specified and congruent with success of a task, stereotypical judgment effects are diminished. That is, when technical characteristics are unknown or insufficient to achieve a task, people seem to rely on stereotypical information (i.e., we replicate previous results with quite similar effect sizes); however, when the technical characteristics are sufficient, people rely more on the technical characteristics than on the "gender" of the robot. As suggested by Fiske [9], people also rely on the objective skills of robots. Besides having replicated results of previous studies that have shown that gender stereotypes affects people's judgments of robots, we have also shown that other indicators consistently affect these judgments and are able to extinguish stereotypical ones. These results are in accordance with predictions of the reflection–construction model [10]. Nevertheless, the equation is not as simple as it seems, as we also found that stereotypes continue to guide people's judgments of male robots, even if the technical characteristics are known and are insufficient to achieve the tasks. It is well-known that judgments do not obey pure rationality but are often biased and based on heuristics (e.g., [13]); thus, the latter result could be explained by representativeness heuristics [14]. Because automotive mechanics is typically a male activity, the male robot could have been judged as representative of the typical car mechanic, so the strong association between the two may have led participants to ignore the actual strength of the robot when making judgments. Further studies are needed to gain a deeper understanding of the asymmetry between judgments based on stereotypes and those based on

rationality and, more generally, to test the generalizability of our results. Our study used a stereotyped male task. It might be interesting to design a study using a stereotyped female task (e.g., caring). If one obtains the same pattern of results, with (a) judgments of a robot based on stereotype when objective indicator is absent (for instance, no information on the ability of a robot to react in an empathetic manner) or insufficient to achieve the tasks (for instance, poor ability to react empathetically) and (b) judgments of a robot based on objective indicator when known and sufficient to achieve the tasks (for instance, excellent ability to react empathetically), it should reinforce the generalizability of our conclusions.

Another limitation and possible extension to our work could be to replicate the study using a real robot. Given that the results indicate that interactions with social robots depend on the nature of interaction with the robot (for instance, real robot vs. simulated robot presented on a screen, see [15]), it would be of interest to use physical robots that interact with participants. In [15], it was found that individuals feel more empathy towards a real robot with which they interact in comparison to an interaction with a simulated robot. Consequently, one could hypothesize that stereotype effects can be boosted when there are interactions with a real social robot because its social presence may favor social-based reactions, perhaps including stereotyping. However, given the study of [6], showing that, in a real interaction paradigm, the stereotype effect seems to be slight if existent, one could expect the same result in a replication of our study using a real robot. Only further theoretical and empirical investigations could help to arbitrate between these two predictions (or even other explanations).

5. Conclusions

In this study we replicated and extended the results found in [4], showing that the effect of human stereotypes on the judgments of robot is not inevitable. We indeed found that participants also rely on technical characteristics when evaluating robots. The effect of gender stereotypes on a robot's ability to succeed in a stereotyped male task was moderated by the strength of the robot. In particular, when available, technical characteristics were used by participants to judge robots with greater accuracy, causing the effect of gender stereotypes to vanish.

Despite the need for future research, we hope that our study will contribute to giving designers of robots the choice between building stereotyped robots and building robots that avoid the perpetuation of human stereotypes, without impacting their potential economic value.

Author Contributions: F. Dufour developed the initial study concept and design of the project. C. Ehrwein Nihan helped further development of the study. Data collection and statistical analyses were performed by F. Dufour and experimenters thanked in acknowledgments. F. Dufour drafted the manuscript, and C. Ehrwein Nihan provided revisions. All authors approved the final version of the manuscript for submission.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Isaac Asimov. "Visit to the World's Fair of 2014." *New York Times*, 16 August 1964. Available online: http://www.nytimes.com/books/97/03/23/lifetimes/asi-v-fair.html (accessed on 27 June 2015).
2. Byron Reeves, and Clifford Nass. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. California and New York: CSLI Publications and Cambridge University Press, 1996.

3. Clifford Nass, Jonathan Steuer, and Ellen R. Tauber. "Computers are Social Actors." In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: CHI'94B, San Jose, CA, USA, 28 April–3 May 2007. Edited by Beth Adelson, Susan Dumais and Judith Olson. New York: ACM Press, 1994, pp. 72–78.

4. Friederike Eyssel, and Frank Hegel. "(S)he's Got the Look: Gender Stereotyping of Robots." *Journal of Applied Social Psychology* 42 (2012): 2213–30. [CrossRef]

5. Benedict Tiong Chee Tay, Taezoon Park, Younbo Jung, Yeow Kee Tan, and Alvin Hong Yee Wong. "When Stereotypes Meet Robots: The Effect of Gender Stereotypes on People's Acceptance of a Security Robot." In *Engineering Psychology and Cognitive Ergonomics. Understanding Human Cognition*, Proceedings of the 10th International Conference on Engineering Psychology and Cognitive Ergonomics, Las Vegas, NV, USA, 21–26 July 2013. Edited by Don Harris. Heidelberg: Springer Berlin, 2013, pp. 261–70.

6. Daniel J. Rea, Yan Wang, and James E. Young. "Check Your Stereotypes at the Door: An Analysis of Gender Typecasts in Social Human-Robot Interaction." In Proceedings of the 7th International Conference on Social Robotics (ICSR), Paris, France, 26–30 October 2015. Edited by Adriana Tapus, Elisabeth André, Jean-Claude Martin, François Ferland and Mehdi Ammi. Cham: Springer International Publishing, 2015, pp. 554–63.

7. Terence Fong, Illah Nourbakhsh, and Kerstin Dautenhahn. "A survey of socially interactive robots." *Robotics and Autonomous Systems* 42 (2003): 143–66. [CrossRef]

8. Alice H. Eagly. *Sex Differences in Social Behavior: A Social Role Interpretation*. Hillsdale: Lawrence Erlbaum, 1987.

9. Susan T. Fiske. "Stereotypes work...but only sometimes: Comment on how to motivate the 'unfinished mind'." *Psychological Inquiry* 3 (1992): 161–62. [CrossRef]

10. Lee Jussim. "Social perception and social reality: A reflection-construction model." *Psychological Review* 98 (1991): 54–73. [CrossRef]

11. Ingo Lütkebohle, Frank Hegel, Simon Schulz, Matthias Hackel, Britta Wrede, Sven Wachsmuth, and Gerhard Sagerer. "The Bielefeld Anthropomorphic Robot Head 'Flobi'." Paper presented at the 2010 IEEE International Conference in Robotics and Automation, Anchorage, AK, USA, 3–8 May 2010, pp. 3384–91.

12. Charles M. Judd, Gary H. McClelland, and Carey S. Ryan. *Data Analysis: A Model Comparison Approach*. New York: Routledge, 2009.

13. Thomas Gilovich, Dale Griffin, and Daniel Kahneman. *Heuristics and Biases: The Psychology of Intuitive Judgment*. New York: Cambridge University Press, 2002.

14. Daniel Kahneman, and Amos Tversky. "Subjective probability: A judgment of representativeness." *Cognitive Psychology* 3 (1972): 430–54. [CrossRef]

15. Stela H. Seo, Denise Geiskkovitch, Masayuki Nakane, Corey King, and James E. Young. "Poor Thing! Would You Feel Sorry for a Simulated Robot? A Comparison of Empathy toward a Physical and a Simulated Robot." Paper presented at the 10th Annual ACM/IEEE International Conference on Human-Robot Interaction, Portland, OR, USA, 2–5 March 2015, pp. 125–32.