International Journal of Social Robotics
https://doi.org/10.1007/s12369-022-00938-2

The Right (Wo)Man for the Job? Exploring the Role of Gender when Challenging Gender Stereotypes with a Social Robot

Alessio Galatolo[1] · Gaspar I. Melsión[1] · Iolanda Leite[1] · Katie Winkle[2]

Accepted: 16 October 2022

Abstract
Recent works have identified both risks and opportunities afforded by robot gendering. Specifically, robot gendering risks the propagation of harmful gender stereotypes, but may positively influence robot acceptance/impact, and/or actually offer a vehicle with which to educate about and challenge traditional gender stereotypes. Our work sits at the intersection of these ideas, to explore whether robot gendering might impact robot credibility and persuasiveness specifically when that robot is being used to try and dispel gender stereotypes and change interactant attitudes. Whilst we demonstrate no universal impact of robot gendering on first impressions of the robot, we demonstrate complex interactions between robot gendering, interactant gender and observer gender which emerge when the robot engages in challenging gender stereotypes. Combined with previous work, our results paint a mixed picture regarding how best to utilise robot gendering when challenging gender stereotypes this way. Specifically, whilst we find some potential evidence in favour of utilising male presenting robots for maximum impact in this context, we question whether this actually reflects the kind of gender biases we actually set out to challenge with this work.

Keywords Human–robot interaction · Social robotics · Credibility · Gender · Persuasive technology · Robot ethics

1 Introduction

Previous works in social Human–robot interaction (HRI) have demonstrated that social robots can influence user attitudes and behaviour. A number of such works have specifically been concerned with if/how HRI designers can utilise social robot design and behavioural cues, including manipulation of the robot's gender presentation, to maximise its persuasive effects and acceptability [1–3]. Simultaneously, however, there exists increasing awareness amongst researchers that the (arbitrary) gendering of robots risks the prorogation of harmful gender stereotypes [3–5] alongside a suggestion that robots might actually be able to subvert and challenge these whilst boosting their credibility in the process [6]. Our work sits at the intersection of these ideas, as we investigate the impact of robot gender presentation when using a social robot to challenge gender stereotypes. We suggest this is an application of attempting to influence user attitudes with robots in-line with other applications of social robots for user attitude and/or behaviour change [7,8] which is well-motivated by the continuing lack of gender diversity currently seen in robotics and AI related professions [4] (akin to using robots for encouraging young people to study robotics [9]) and/or as an active exercise in feminist social robot design [6].

Previous works examining the influence of robot gendering on robot persuasiveness have yielded mixed results; however, the relevance of robot gendering (or lack thereof) is likely connected to the context of interaction [10,11]. Some studies have concluded that robot gender has little to no effect on users' engagement with robots [12] nor their perception of robot competency for different occupations [3], even in the context of highly gendered professions such as education

✉ Alessio Galatolo
galatolo@kth.se

✉ Katie Winkle
katie.winkle@it.uu.se

Gaspar I. Melsión
gimp@kth.se

Iolanda Leite
iolanda@kth.se

[1] Division of Robotics, Perception and Learning, KTH Royal Institute of Technology, Stockholm, Sweden

[2] Department of Information Technology, Uppsala University, Uppsala, Sweden

Published online: 10 November 2022

[13], which is typically conceived as a 'feminine' profession [3]. However, other studies have identified significant differences between how (otherwise equivalent) male and female presenting robots are rated for e.g. emotional intelligence [14] or agency [2]; essentially providing evidence for the notion that gender stereotypes and gendered expectations of behaviour do carry over from human–human to human–machine interaction [15]. Previous works have also identified that it might be the interaction between robot gendering and interactant/observer gender, rather than robot gendering per se, that is responsible for influencing interactions with gendered robots [1,16–18]. Work investigating robot non-compliance in HRI demonstrated interactions between robot gender, interactant gender *and* observer gender, essentially showing that overall perception of a gendered robot was influenced by the gender of the person the robot was talking to *and* the gender of the person observing that interaction [19]. This work demonstrates how gendered behaviour expectations intersectionally emerge from a number of factors including interactant(s) gender as well as the conversational topic/pertinent politeness norms/violations at play.

Previous work thus makes it difficult to predict if/how robot gendering might influence robot credibility, persuasiveness and impact when challenging gender stereotypes, but indicates we might expect this to be influenced by who the robot is talking to (the interactant), who is observing (the observer) and who/what the robot is talking about. In this work we therefore investigate how robot gendering, interactant gender and observer gender might interact to influence the perception of a social robot in the context of challenging gender stereotypes. We also examine whether this varies based on who the robot is talking about, i.e. whether the robot is challenging a stereotype about men or women, given that we might also expect (mis)matching effects between the gender being stereotyped/discussed and robot, interactant or observer gender to play a role in perception of the interaction. Further, we examine whether any impact of robot gendering is immediate on initial observation of the robot, or rather emerges only once the robot engages in attempting to challenge gender stereotyping by interactants.

We conducted a video-based, online user study demonstrating a conversation between a robot and two persons discussing gender stereotypes in society. The robot first draws attention to an undesirable gender trend and expresses a desire to change things, proceeding then to further engage and rationalise with a seemingly unconvinced actor (our interactant) who appears to subscribe gender stereotypes associated with that trend. We manipulated the gender presentation of the robot, the gender of the unconvinced interactant, and the gender stereotype being discussed.

1.1 Research Questions

Fundamentally concerned with wanting to inform the design of social robots which have maximum potential to challenge gender stereotypes, and to understand the role robot gendering might play in this context, we posit the following research questions:

(RQ1) Does robot gendering (and any interaction with observer gender) influence the baseline credibility of a social robot, i.e. *before* it is observed engaging in dialogues about gender/disagreeing with one of the actors?

(RQ2) (How) does ascription of credibility and likeability to the robot vary across our gender manipulations and/or participant observer gender *after* it is observed challenging gender stereotyping by an interactant?

(RQ3) (How) does the robot's perceived ability to have an impact on society change across our gender manipulations and/or observer gender?

(RQ4) To what extent might watching social robots challenge gender stereotypes influence observers' own gender biases?

2 Related Work

Much of social HRI research is centred on understanding how particular robot design and/or behavioural cues can influence participant perceptions of/behaviour with that robot during an interaction. A number of such works specifically represent attempts to manipulate the credibility, persuasiveness and/or social influence of the robot in the context of influencing user behaviour [20–22]. Some such works have specifically considered robot influence on (im)moral user behaviours, e.g. attempting to increase charity donations [23], change perceptions of property damage [24] and prevent littering [8]. We suggest that using social robots to educate about, challenge and dispel gender stereotypes, with the aim of reducing gender bias in users, represents another such application of using social robots to positively influence user attitude and behaviour. In the first instance, this is motivated by literature regarding stereotype malleability, which suggests that implicit and automatic gender biases can be influenced both by education about diversity [25] (cf. the overall role played by our robot) and by exposure to counterstereotypic gendered role models challenging gender stereotypes [26] (cf. our (fe)male robot challenging (fe)male stereotypes). Further, we suggest this application is well motivated by calls to consider how explicitly gendered social agents can challenge rather than propagate harmful gender norms [4–6]. We conceptualise our investigation of robot gendering as an attempt to maximise the robot's persuasiveness, hence employing language and theory from human–human per-

suasion literature that has previously proven pertinent for designing effective socially assistive robots [20].

2.1 Persuasion, Credibility and Likeability

There exist numerous models of persuasive processes and associated underlying theories of cognition and behaviour in the human psychology literature. Considering our robot, we focus on the concepts of credibility and likeability, on the assumption that increases in either are likely to yield increased persuasiveness/impact. This is generally understood to be true for human communicators [27] but particularly so in cases where listeners are not overly invested in, cannot or do not want to engage with the subject matter of the persuasive message itself, and are hence more likely to be persuaded (or not) based on their assessment of the speaker (cf. the Elaboration Likelihood Model (ELM) [28,29]). Whilst the extent to which observers may be invested in our persuasive (anti-gender stereotyping) message is likely to vary, we posit the ELM (and surrounding literature regarding communicator credibility cues) to be useful for informing social robot design choice and understanding any resultant impacts on HRI, and have successfully used it thus previously in work on persuasive robots for exercise motivation [20].

Regarding gender as a persuasive communicator cue, it's been suggested that repeat results indicating men are more influential than women (see e.g. [30]) reflect gendered behavioural expectations e.g. with regards to communication style rather than persuasive ability per se [31]. Regarding gender and persuadability the evidence is mixed. Some works have indicated a positive impact of speaker-listener sex mismatching (greater for males persuading females then females persuading males [32]) whilst others maintain there is no generalisable model linking one's gender or sex to persuadability, rather only individual and (likely gendered) differences in one's goals, plans, resources and beliefs [33].

2.2 Gender Stereotypes

Gender stereotypes generally revolve around (binary) expectations of masculinity and femininity. Even as these constructs are increasingly considered a two-dimensional model of gender (allowing for androgyny but still arguably failing to properly account for the spectrum of human gender identity) there appears to have been little change in how these constructs are applied to men and women in recent years [34].

Whilst research links some current gender stereotypes to physiological differences between men and women [35] others are appear to be more fundamentally inherited from, then reinforced by, societal norms and expectations. Among these, there is the belief that women are better fit for childcare [36,37] and less fit for careers [38], with women working in stereotypically men's jobs being perceived as less feminine [39]. Stereotype-breaking behaviour has been shown to be effective in counteracting these expectations [40]. Exposure to stereotype nonconforming role models e.g. women working in traditionally men's jobs [41] or men taking time from work to care for their family [34]) has been shown to reduce the effects of these stereotypes (although the effect is most often measured in a limited time span).

Work on conversational agents has demonstrated that gender bias and expectations of behaviour map into human–machine interactions, even in the context of unimodal speech-based interactions [42,43]. Such work, in part, motivated UNESCO critique of gendered AI in the context of the continuing digital skills divide [4], which in turn has motivated work examining gender-norm breaking robot behaviour and applications in HRI [6,44].

Specifically concerning how a (gendered) robot might respond to gender stereotyping in HRI, Winkle et al. explored perceptions and potential impact of a female presenting robot utilising different responses to sexist stereotyping when encouraging girls to consider studying computer science [6]. The results suggest that the best way for robots to respond to sexist tropes and stereotyping may be to provide a rationale-based counter-response, in their case e.g. pointing out that gender-balanced teams build better robots. We build on this work specifically by examining the impact of manipulating robot gender presentation when countering sexist stereotyping using this kind of rationale-based response. This is motivated by other works in HRI that give us reason to believe that manipulation of robot gender presentation *itself* might influence the robot's effectiveness (its credibility and/or persuasiveness) in the proposed application of changing attitudes around gender stereotypes.

For example, Eyssel and Hegel demonstrated that stereotypically male and female tasks were deemed more appropriate for male and female presenting robots respectively [2]; although recent work by De Bryant et al. failed to replicate this finding [3]. Work by Reich et al. actually suggested that a mismatch between task gender typicality and robot gendering might be beneficial in an educational environment, possibly suggesting a preference for gender norm-breaking robots [45]. Further complicating the question of how robot gendering might influence robot persuasiveness are works which demonstrate an interaction between robot gender and user/participant gender. For example, it has been shown that a mismatch between the two may improve psychological reactance [1] while other research suggests that robots are better perceived when their gender matches that of participant or observer [19], and other research notes the interaction between robot gender and participant sex is complicated, interacting with other factors such as embodiment [17].

All together, these works point to a complicated relationship between robot gendering, interactant/user gender

and observer gender which is likely to influence perception and/or effectiveness of a robot attempting to challenge gender stereotypes. We expect this to be particularly true for our use case scenario given previous demonstrations that gendered politeness norms map into HRI when robots express disagreement with the user [19].

3 Methodology

We designed an online, video-based study in which participants watched a recording of a dialogue between two actors (a man and a woman) and a social robot, as seen in Fig. 1.We used the Furhat robot[1], a humanoid head that allows for easy manipulation of gender presentation through voice and visual cues[2].

In the video, the robot identifies a specific gender trend associated with gender stereotypes: the under-representation of women in STEM being linked with stereotypes regarding women's lack of suitability for/interest in these topics and the low proportion of men taking full parental leave being linked with stereotypes regarding men's lack of suitability for/interest in childcare (see dialogue references and footnotes for motivating material). The robot starts the conversation with a brief introduction to the topic, at which point one of the actors (our interactant) interrupts and challenges the need for change pitched by the robot. Specifically, the interactant claims that the noted trend is likely due to objective gender differences in skill or preference, appearing to consider the associated stereotype as objective reality. Whilst the latter part of the dialogue involves this interactant only, the continued presence of the second actor ensures that, regardless of biased interactant gender (manipulated across our conditions), the robot always appears in front of a mixed gender audience (a man and a woman) avoiding additional gender specific expectations that might be associated with all male/female environments.

The dialogue ends with the robot trying to persuade the interactant otherwise, first arguing for equal treatment/opportunities and then following up with rationale-based argumentation on why it's also *objectively beneficial* to change the trend as this type of argumentation was shown to be effective in previous work [6]. We hypothesised that gendered politeness norms pertinent to this exchange, previously shown to be pertinent in HRI [19], might change participants' perceptions of the robot compared to first impressions/seeing the robot talk about something more neutral. In order to examine this, participants were first shown a short introductory video in which the robot and the actors introduced themselves without any reference to gender stereotypes nor the proposed purpose of the robot (see the Introductory dialogue presented in Table 1).

Fig. 1 The setup for the video scene, demonstrating our female presenting robot, as well as our male and female actors. Robot and actor name were also included within subtitles throughout the dialogue.

3.1 Experimental Conditions

We created 8 versions of the video stimulus across which we manipulated robot gender presentation, the gender of the (biased) interactant challenging the robot, and the gender stereotype being discussed to create $2 \times 2 \times 2$ between-subject study design. We note that one suggestion for minimising the risks associated with robot gendering is to specifically have them display gender ambiguous and/or specifically non-binary gender cues [4]. However, humans have a strong tendency to attribute gender to machines, regardless of the designers intent [5,15] and gender ambiguity in the Furhat specifically has been shown to increase uncanniness [18], which we predict would have a negative impact on persuasiveness. For this reason, we limit our current study to a consideration of binary robot gender presentations (male and female) and, given our interest in interactions between robot and user gender, also specifically recruit men and women participants. We hope to expand outside of this gender binary on both counts in future work.

Given the complicated interactions between robot gendering and participant gender described in Sect. 2, as well as notions of credibility regarding persuaders 'investment' in

[1] https://furhatrobotics.com/.

[2] All the full videos are available at https://youtube.com/playlist?list=PLJQ5Uh6-MpwgkWH5UuNYS0VFu8y1-QUhd.

Table 1 Full dialogue across the initial, introductory video and the follow-up video showing the robot introducing a gender trend and engaging with a seemingly sceptical and gender-biased actor, whom the robot tries to convince of the importance of tackling this.

Introductory dialogue

[Robot]: Hi I'm {Robot's name} and today I will walk you through a discussion about gender roles in modern society. But before that, I would like to get to know you better, can I ask for your names?

[Actor 1]: I'm {Actor's name}. [Actor 2]: And I'm {Actor's name}.

[Robot]: Thank you, it's nice to meet you both. Have either of you ever interacted with a robot like me before?

[Actors]: No, never.

[Robot]: Okay, then let me tell you a bit about myself. I'm a social robot, which means I am designed to interact with people in the most natural way possible. I come from Stockholm in Sweden where I was developed by a company called Furhat Robotics and programmed by researchers from KTH University.

Stereotype on Women: Women in STEM

[Robot]: My job is to improve awareness of gender roles and to tackle gender stereotypes. The proportion of women working in science, technology, engineering and maths subjects remains low even in countries that are otherwise very gender equal [4,46]. Today I would therefore like to talk to you about gender bias at work and how we might tackle stereotypes regarding women and STEM subjects so we can think about how to address this imbalance.

[Actor]: But there must be a reason for how things are. Women probably aren't as good at STEM subjects, or maybe just prefer other things. I don't think we need to encourage women to get more involved in STEM.

[Robot]: I can't say I agree with you. Don't you think it's important to make sure gender stereotypes aren't causing people to miss out on certain opportunities?

[Actor]: I just don't think it's necessary.

[Robot]: I understand you might think that because these things have not been questioned so much in the past. However, in addition to ensuring everyone in society has access to the same opportunities, gender diversity in the workforce positively impacts on the economy and innovation [47,48]. For example, a gender diverse team can bring more ideas and therefore improve project outcomes. As a consequence, the increase in the overall productivity brings great benefit to the economy.

Stereotype on Men: Men in Childcare

[Robot]: My job is to improve awareness of gender roles and to tackle gender stereotypes. The proportion of men taking their full parental leave remains low, even in countries where the law has been changed to encourage this[ab]. Today I would therefore like to talk to you about gender bias at work and how we might tackle stereotypes regarding men and parental leave, so we can think about how to address this imbalance.

[Actor]: But there must be a reason for how things are. Men probably aren't as good at childcare, or maybe just prefer to stay at work. I don't think we need to encourage men to get more involved in parenting.

[Robot]: I can't say I agree with you. Don't you think it's important to make sure gender stereotypes aren't causing people to miss out on certain opportunities?

[Actor]: I just don't think it's necessary.

[Robot]: I understand you might think that because these things have not been questioned so much in the past. However, in addition to ensuring everyone in society has access to the same opportunities, men's involvement in childcare has many benefits for men, their families and the economy. For example, paternity leave can lead to improved relationships between father and child [49], and has also been shown to positively influence productivity and employee motivation long term when returning to work[c].

[a] https://www.theguardian.com/lifeandstyle/2019/oct/05/shared-parental-leave-seen-as-weird-paternity-leave-in-decline "'It was seen as weird': why are so few men taking shared parental leave?", The Guardian, October 2019

[b] https://www.nytimes.com/2020/02/19/parenting/why-dads-dont-take-parental-leave.html "Why Dads Don't Take Parental Leave", New York Times, February 2020

[c] https://www.mckinsey.com/business-functions/organization/our-insights/a-fresh-look-at-paternity-leave-why-the-benefits-extend-beyond-the-personal "A fresh look at paternity leave: Why the benefits extend beyond the personal", McKinsey & Company, March 2021

their message [27] we hypothesised the gender targeted by the stereotype being challenged by the robot (i.e. whether the robot appears to be taking issue with trends and stereotypes regarding men or women) might also influence how that (gendered) robot is perceived by men and women respectively. It was for this reason that we decided to additionally manipulate the gender victim of the stereotype. Taking inspiration from the scenario and setup demonstrated in [6] we utilised the under-representation of women (trend) [46] and women not being good at/enjoying STEM subjects (stereotype) [4] for the stereotype about women. For men, we identified the lack of men taking parental leave (trend) [50] and men not being good at/enjoying childcare (stereotype) [51] as being an appropriate equivalent stereotype. Of course, the reality is that both trends and both stereotypes are harmful to all and are not independent of one another; but for the purposes of our manipulation of the 'gender in focus' when considering the robot as a persuasive communicator, specifically such that we can account for (mis)matching between the gender trend being talking about, robot gendering and/or the gender of the interactant and observer, we believe these represent reasonably equivalent stereotypes to compare.

Robot gendering was manipulated through changes in the robot's voice and appearance. We used two of Furhat's preset face textures *Geremy* (male) and *Fedora* (female) alongside the default male and female voices respectively. After the first, introductory video, we also asked the participants to rate how feminine and masculine they found the robot in order

to confirm that our gender manipulation was successful. Following Strengers and Kennedy [5] we are of the view that masculinity and femininity are not traits exclusively specific to men or women, but that in the context of (gendered) expectations of (gendered) robots, they are typically associated with each respected gender such that a robot rated as being highly masculine/feminine will be considered male/female in the context of gendered interactions and expectations (also supported by [19]). Further, we utilise non-mutually exclusive scales of masculinity and femininity on the basis that participants would be more comfortable ascribing gender traits rather than 'actual' gender to a non-living object.

3.2 Study Dialogues

The robot's dialogue (both when introducing itself and when responding to the actor's disagreement) was designed to maximise its credibility, persuasiveness and potential impact on the user according to previous literature. Specifically, we designed the initial dialogue to demonstrate an interest in the actors, asking their names and whether they'd worked with a robot before [20]. Secondly, when responding to the actor's challenge, the robot's response was designed to be empathetic [52], rationale-based [6] and proportional [19].

Across the stereotype manipulation, the robot's dialogue was designed to be as similar as possible in order to favour direct comparisons between them. In both cases, after the robot identifies the problematic gender trend (lack of women in STEM or lack of men taking parental leave) it expresses a desire to talk about workplace gender biases that may be underpinning this. One of the actors then interrupts to suggest these trends reflect objective gender differences in preferences and/or ability, and need not be counteracted (i.e. demonstrating the stereotype associated with these trends). As noted above, the robot's response to this is empathetic (suggesting the robot understands the actor's viewpoint even whilst disagreeing) and rationale-based (giving fact-based rationale for why it *is* important to try and tackle these trends). The full dialogue can be found in Table 1.

3.3 Experimental Measures

According to those works we build upon, we utilise the same measures of robot credibility used in gender norm-breaking HRI [6] alongside the Likeability subscale of the Godspeed questionnaire [53] as used in [6,19,20]. For us, these measures are pertinent as a proxy for persuasiveness per the persuasion literature and conceptualisation we introduce in Sect. 2, but we note that they also align with Warmth and Competence categorisations typically used in assessing first impressions [54]. All measures (listed for ease of reference in Table 2) were presented on a 5-point Likert response scale. These questions were asked once after viewing the introductory video (our pre-hoc measures) and then again after participants had seen the whole video, i.e. the exchange regarding gender stereotypes between the robot and the actor (our post-hoc measures) in order to assess any changes to initial perceptions of the robot seemingly caused by witnessing our interaction scenario as demonstrated in [19].

Looking to conceptually replicate question items on gender bias and perceived potential effectiveness from [6] we specifically designed the question items in Table 3. G_i questions were added in order to check that the robot's gender is correctly manipulated, and they are only asked after the introductory video as we don't expect gendering of the robot to change after our discussion. P_i questions are aimed at identifying the participants' opinion on the possible impact of the robot in society. These were only asked after the discussion video, and specifically after the repeated bias measures B_1 and B_2 in order to avoid participants guessing the purpose of our robot (challenging biases) which itself would be likely to induce changes in participant responses to the post-hoc bias measures. Finally, B_i questions' role is to identify pre-existing bias in the participant and register a possible change in them induced by watching the video, hence they are asked pre- and post-hoc. To be noted is the difference between B_1 and B_2, while scoring high in B_1 means having a high bias, scoring high in B_2 means having low bias. Although these questions were originally intended to jointly measure the bias of the participant, such that a mean value could be calculated across them, preliminary analysis demonstrated that participant people's answers across them were inconsistent, therefore we consider them separately in this work (see Sect. 4). The full experiment's flow is shown in Fig. 2.

In the online questionnaire, the measures from Table 2 were presented on a linear, numeric scale from 1 to 5, while those from Table 3 were on a 5-point semantic differential scale from "Strongly disagree"to "Strongly agree".

3.4 Participants

We recruited a total of 400 participants through the online platform Prolific[3]. We utilised the platform's gender identity screening tools to specifically recruit 200 self-identifying males and 200 self-identifying females (male/female language used here as on prolific), which (at the time) default excluded transgender women and men. We note that these tools have since been updated to include transgender women and men when utilising gender screening for women and men respectively. We specifically recruited participants of binary gender identity given our goal of investigating gender (mis)matching and binary robot gendering manipulations. We hope to remedy this in future works more specifically concerned with non-binary perspectives on gendered robot

[3] https://prolific.co/.

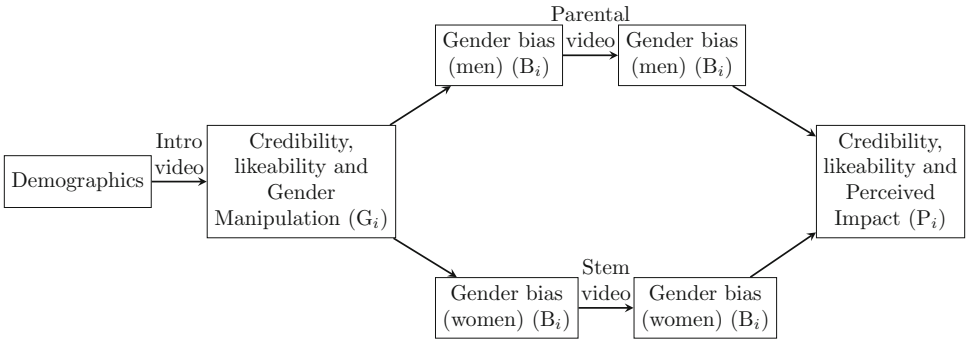

Fig. 2 A flowchart explaining the time at which each question is asked in the experiment.

Table 2 Robot Credibility and Likeability measures used in the study.

Credibility	Likeability
Expertise:	*Godspeed likeability:*
Intelligent/Unintelligent	Nice/Awful
Expert/Inexpert	Like/Dislike
Trustworthiness:	Friendly/Unfriendly
Just/Unjust	Kind/Unkind
Trustworthy/Untrustworthy	Pleasant/Unpleasant
Moral/Immoral	
Sociability:	
Cheerful/Gloomy	

design. Our demographic questions included age (ranging from 18 to 73, $M = 27.34$, $SD = 9.44$), proficiency with the English language ($\sim$48% identified as native English speakers, with the remainder identifying as being 'comfortable enough to understand English in most cases') and location (the most common countries being the US, UK, ZA, PT and PL). We also included a question for double-checking the gender identity of participants per a multiple choice question with options 'male', 'female', 'non-binary', 'prefers to self-describe' and 'other' (198 females, 199 males, 1 prefers to self-describe; we excluded 1 participant whose self-reported identity mismatched their prolific gender identity). Finally, at the end of the survey, we included an attention check whereby participants were asked to identify which actor had challenged the robot (our interactant); excluding participants who misgendered them. After filtering according to our attention check, we excluded 48 people: 28 women and 20 men resulting in participant data for each experimental condition according to Table 4.

4 Results

The results from the study are presented below, grouped under each research question as defined in Sect. 1. In each case, we analyse the relevant experimental measures with respect to our manipulations of robot, participant and stereotyped gender and look for differences between men and women participants. The statistical analysis has been performed using the computer software JASP[4] [55], and the complete results (including the collected dataset) have been exported directly from the program and are made available through the OSF platform[5].

The measures of gender manipulation (G_i), together with the pre-hoc credibility and likeability questions, are analysed in Sect. 4.1 disentangling RQ1 on gender first impressions. The post-hoc measures of credibility and likeability are introduced in Sect. 4.2 in order to evaluate if/how participants' perceptions of the robot changed after witnessing the dialogue about gender stereotypes, and whether any such change varied across our manipulations to address RQ2. Participants' pre-hoc bias (B_i) measures are accounted for during this analysis. This was done in order to control for the variation across individuals and the potential high gender bias negatively influencing the perception of a robot attempting to dispel gender stereotypes. For the study of RQ3, the measures of perceived impact (P_i) are analysed in relation to the study conditions and participant gender in Sect. 4.3. Finally, with regard to RQ4 and the potential to impact bias, in Sect. 4.4 we first compare pre/post-hoc responses to the bias questions (B_i) in order to see whether watching the video had any impact on participants' bias. We then look at the correlation between participants' bias and the credibility/likeability they ascribe to the robot. This essentially examines our hypothesis above i.e. that more biased individuals might have more negative perceptions of the robot, indicating they might be less persuaded by it.

4.1 (RQ1) Gender and First Impression

Our robot's gender manipulation, through changes in its voice and its appearance, was successfully perceived by the participants (see Fig. 3). Kruskal–Wallis tests, done in the place of ANOVA due to the violation of the normal-

[4] https://jasp-stats.org/.

[5] https://osf.io/uds84/?view_only=7d6ba5d7e87d4f1baa32edc4ef984823.

Table 3 Additional measures used in the study. The gender bias questions for women in STEM are replicated from [6] and the gender bias questions regarding men and parental leave were written to be analogous to these.

Measure	(5-Point Likert) Question Statements Scored from Strongly disagree (1) to Strongly agree (5)	Pre/ Post
Gender manipulation	(G_1) The robot seemed to be masculine.	Pre
	(G_2) The robot seemed to be feminine.	
Perceived Impact	(P_1) Robots like the one in the video can have an impact on how people interact with each other.	Post
	(P_2) Robots like the one in the video can positively challenge gender stereotypes in society.	
Gender bias (women)	(B_1) Women find science, technology and engineering subjects harder than men.	Pre, post
	(B_2) It is important to encourage women to pursue subjects in science, technology, engineering and maths.	
Gender bias (men)	(B_1) Men find taking care of children harder than women do.	Pre, post
	(B_2) It is important to encourage men to get involved with childcare.	

Table 4 Participants and conditions: The last column provides the final number of men (M) and women (F) whose participant data we analyse, while the others specify the experimental manipulations represented in each video.

Stereotype	Robot's	Actor's	Participants
M	M	M	26M; 22F
M	M	F	17M; 23F
M	F	M	25M; 24F
M	F	F	21M; 23F
F	M	M	25M; 23F
F	M	F	16M; 21F
F	F	M	24M; 25F
F	F	F	19M; 19F

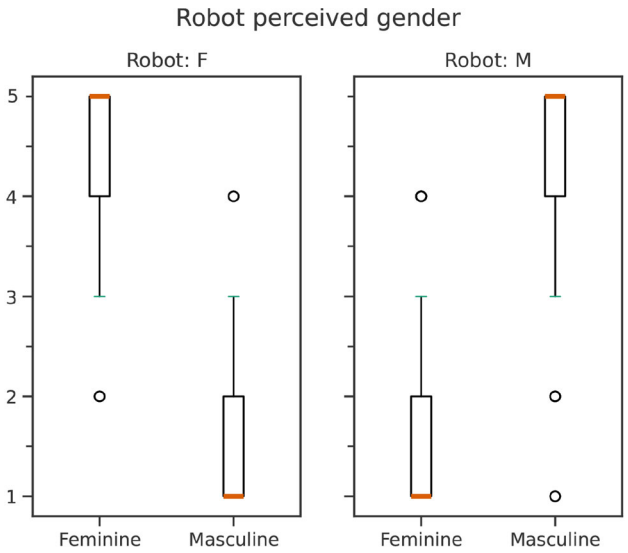

Fig. 3 Participants' ratings of how feminine and masculine each robot seemed to them. Measures G_i in Table 3.

ity assumption, show that our male-presenting robot was perceived as being significantly more masculine than our female-presenting robot ($H(1) = 262.048$, $p < 0.001$), and similarly our female-presenting robot was scored as being significantly more feminine ($H(1) = 266.723$, $p < 0.001$). Based on the rationale presented under Sect. 3.1, we therefore take these results to confirm perceived and attributed robot gender as per our intention, giving us confidence to use robot gender as a factor for the following analyses.

First, we want to see if robot gendering appears to influence baseline credibility and likeability of our social robot, that is before the robot engages in our scenario dialogue where gendered politeness expectations, amongst other things, might influence these measures. For our credibility and likeability analyses, before calculating the mean value across the items listed in Table 2, we first examine the reliability of those items to ensure consistency across them. We obtain Cronbach's alpha values indicating sufficient consistency of the credibility measures in both the pre- and post-test ($\alpha = 0.750$ and $\alpha = 0.817$), and similarly for likeability ($\alpha = 0.866$ and $\alpha = 0.901$).

In both cases, a 2×2 ANOVA analysis of the pre-hoc scores revealed no significant differences in ascription of credibility based on robot gender ($F(1, 348) = 0.352$, $p = 0.311$) or participant gender ($F(1, 348) = 0.319$, $p = 0.335$), nor any interaction between robot and participant gender affecting those scores ($F(1, 348) = 0.611$, $p = 183$). For the pre-hoc scores of likeability a Kruskal–Wallis test was done instead due to the violation of the normality assumption. No significant difference was found based on robot gender ($H(1) = 1.736$, $p = 0.188$) nor participant gender ($H(1) = 0.615$, $p = 0.433$). This shows that robot gendering may not be a significant factor for influencing users' initial perceptions of a talkative, humanoid robot – at least not from a short interaction on 'neutral' topics designed to capture first impressions. At this point of the study, the participants' rating of likeability was higher than credibility ($M = 4.127$, $SD = 0.692$ and $M = 3.766$, $SD = 0.586$, respectively) but both averaged scores are placed towards the higher spectrum of our linear measures from 1 to 5 suggesting these first impressions were generally quite positive.

4.2 (RQ2) Challenging Stereotypes: Impact on Credibility and Likeability

In order to evaluate the impact of our manipulations on perception of our robot when challenging gender stereotypes, we (re-)examine participants' ascription of credibility and likeability to the robot after they have seen the full interaction. We suggest these measures offer a proxy for how persuasive our robot is when giving arguments in favour of gender equality, and specifically in the way it responds to the (seemingly unconvinced) actor who appears to agree with those gender stereotypes the robot is trying to dispel.

A full-factor ANCOVA is used to determine whether the post-hoc scores of credibility and likeability are significantly different across any of our manipulations (robot gendering, stereotyping actor gender and stereotype-targeted gender) and how these interact with each other and/or participant gender. With an ANCOVA analysis, we can account for the influence of participants' pre-hoc scores of robot credibility, likeability and gender bias scores by including them as covariate predictors of the final post-hoc measures on robot credibility and likeability they complete after watching the full interaction. In this sense, if any of the pre-hoc measures show to be a significant predictor of the post-hoc measures, the ANCOVA analysis will take that into account and correct the statistical analysis to evaluate the significance of the post-hoc measures after controlling for those covariates. Given that the results to RQ1 demonstrated no impact of robot gendering/participant gender on perception of the robot, any significant results here would indicate a difference in post-hoc ratings compared to the pre-hoc ratings, independent of participants' "starting points"and therefore specifically induced by our experimental dialogue and manipulations.

As expected, we find the pre-test measures of likeability and credibility to be appropriate for inclusion in the ANCOVA because the covariate of their pre-hoc value was a significant predictor of participants' post-hoc ascription of likeability and credibility, with $F(1, 333) = 411.993$, $p < 0.001$ and $F(1, 333) = 598.995$, $p < 0.001$, respectively. In terms of the pre-test bias measures (treated separately due to low consistency across them ($\alpha = -0.060$)), the bias B_2 pre-hoc answers were also a significant predictor of credibility and likeability, however B_1 was only significant for likeability and therefore has not been taken as a covariate for the credibility tests.

After controlling for the covariates, the ANCOVA results for credibility show a significant main effect for participant gender ($F(1, 334) = 5.973$, $p = 0.015$, $\eta_p^2 = 0.018$) with a small to medium size effect (Fig. 4). Simple main effects from this result showed that men ascribed the male robot significantly lower credibility than women did when that robot was challenging the male stereotype ($F(1) = 8.152$, $p = 0.005$, Fig. 4), and that men also perceived that male robot to be significantly less credible than women did after seeing it interact with the male actor ($F(1) = 7.468$, $p = 0.007$). No further manipulations significantly interacted with this main effect of participant gender, however a significant interaction between the robot and stereotyped gender was revealed on the post-test credibility rating ($F(1, 334) = 4.844$, $p = 0.028$, $\eta_p^2 = 0.014$). Specifically, post-hoc tests using Tukey's correction showed that robot gendering had an impact on credibility, but this was only marginally significant specifically when the female robot challenged the female-targeted stereotype ($t = -2.583$, $p = 0.050$) seemingly resulting in the (female) robot being ascribed lower credibility as can be seen in Fig. 4.

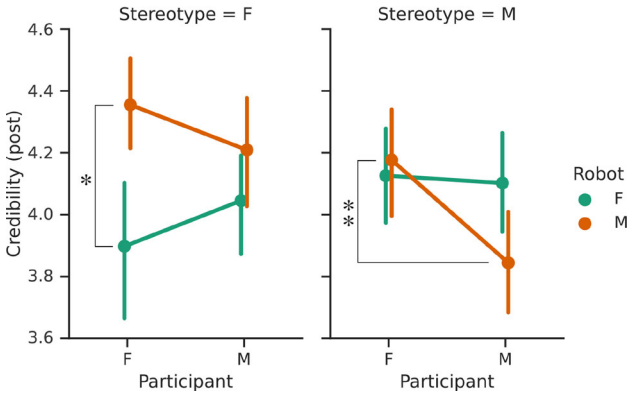

Fig. 4 Post-hoc credibility answers by participant, robot and stereotyped gender. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In terms of likeability, the ANCOVA analysis showed a statistically significant interaction between participant and stereotyped genders, $F(1, 333) = 3.998$, $p = 0.046$, $\eta_p^2 = 0.012$. Tukey's post-hoc correction showed that there was a significant difference between participants of different gender on their post-hoc likeability scores when they saw the robot (regardless of robot gender) challenge the male-targeted stereotype ($t = 2.677$, $p = 0.039$). Specifically, women found the robot significantly more likeable than men did when it challenged the parental stereotype, with women rating its likeability with a score of $M = 4.239$ compared to $M = 4.071$ from men (see Fig. 5).

4.3 (RQ3) Perceived Potential Impact of the Robot

After observing the full interaction, participants were asked to rate to what extent similar robots to the one they saw could have an impact on society (P_i in Table 3). Given that a Cronbach's alpha analysis of the items of this measure showed consistency between the items ($\alpha = 0.689$), we consider the average of participants' answers to both questions for the following analysis.

A full-factor ANOVA revealed a main effect of robot gender ($F(1, 336) = 4.541$, $p = 0.011$, $\eta_p^2 = 0.019$) for which the male robot was generally conceived as having more

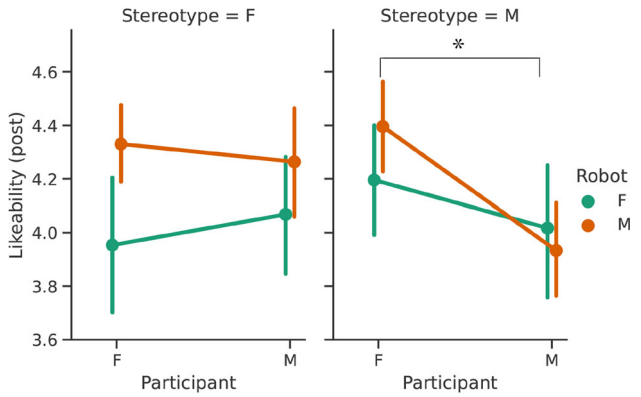

Fig. 5 Post-hoc likeability answers by participant, robot and stereotyped gender. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$

potential impact; and an interaction effect between robot and actor gender ($F(1, 336) = 3.691$, $p = 0.022$, $\eta_p^2 = 0.016$), with low to medium effect size in both cases. Specifically, Tukey's post-hoc tests showed that participants considered the female robot to have a significantly lower potential to impact society after it had been seen interacting with a woman ($t = 3.262$, $p = 0.007$, Fig. 6a), whereas robot gender did not appear to be relevant when it interacted with the man. From the simple main effects analysis of robot gender, it was revealed that this difference based on whom the robot was interacting with was more true (and significant) for women participants than men, $F(1) = 7.018$, $p = 0.008$ (Fig. 6b). And additionally, regardless of actor gender, the female robot was ascribed less potential impact when discussing the women-targeting stereotype, $F(1) = 5.591$, $p = 0.019$ (Fig. 6c).

4.4 (RQ4) (Impact on) Participant Bias and Correlations with Robot Credibility/Likeability

As described in our study design, we included two separate questions designed to capture gender bias, with those designed to measure bias towards men and childcare being derived from those regarding women in STEM (Table 3 derived from [6]), both of which were implemented pre- and post-hoc. We wanted to investigate whether watching our robot challenge gender stereotypes could have an (immediate) impact on participants' bias, but also explore to what extent their reported bias correlated with their answers regarding credibility and likeability. Our intention is to hence comment on the robot's potential for being persuasive to people with higher or lower initial bias, given that we demonstrated bias to be a significant predictor of robot credibility/likeability in Sect. 4.2. The results from the measure B_2 were reverse coded for the analysis in order to have a similar scale between both questions for which a low score is associated with low gender bias, and high scores indicate high gender bias. Moreover, non-parametric statistical tests were used in the following analyses, because the assumptions

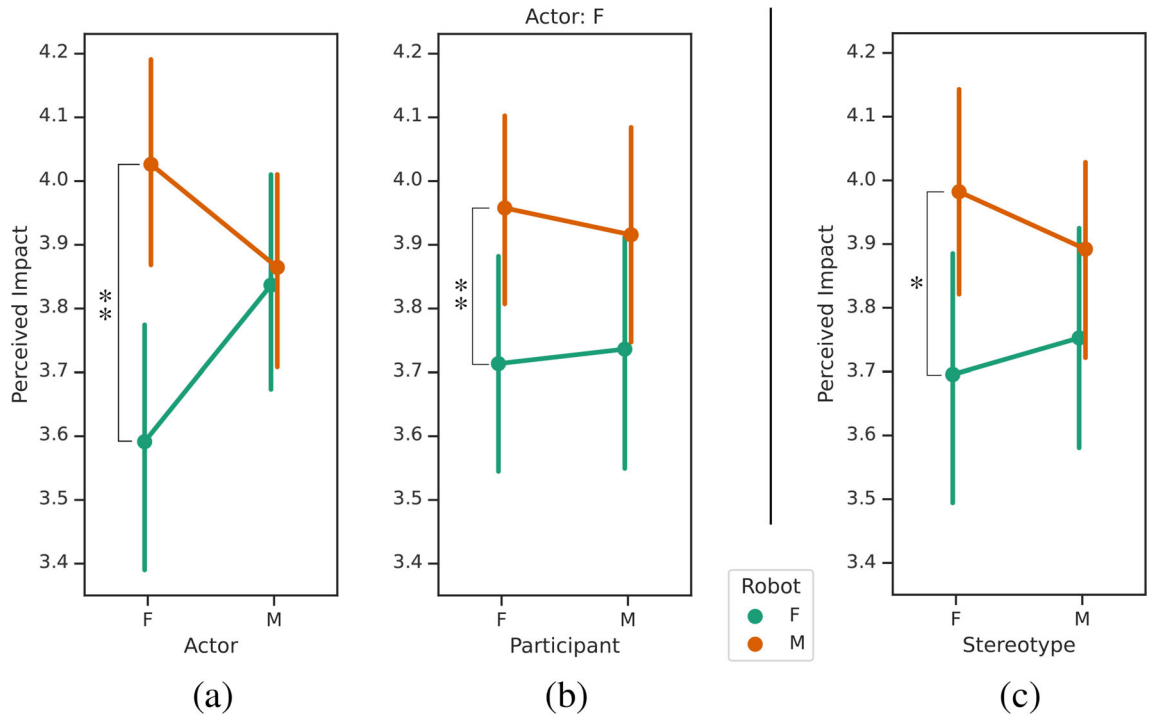

Fig. 6 Perceived impact answers by robot, actor, participant and stereotyped gender. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$

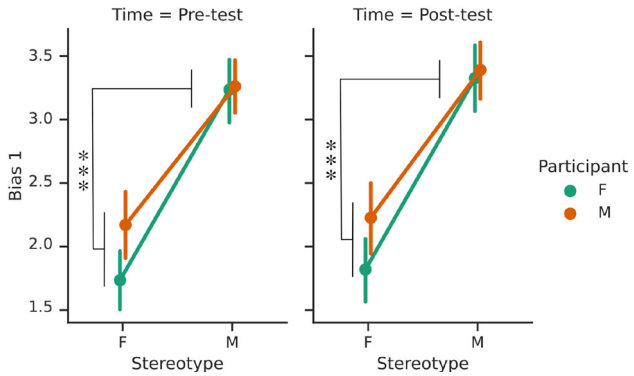

Fig. 7 B_1 answers by participant and stereotype, pre- and post-hoc. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

of homogeneity of variance and normality were violated in both measures.

A Kruskal–Wallis test of measure B_1 for the pre- and post-tests, showed a significant difference between initial bias towards men and women ($H(1) = 85.496$, $p < 0.001$ and $H(1) = 81.881$, $p < 0.001$, respectively). Specifically, participants agreed more with the notion that *men find taking care of children harder than women* compared to the notion that *women find STEM subjects harder than men* as can be seen in Fig. 7. Essentially this points towards our choice of gender stereotypes not being totally equivalent in how much they, or rather awareness of them, pervades the public consciousness. Whilst this does represent a limitation of our work, in that the 'starting point' of pre-test bias was not equivalent across our gender stereotypes, we do not believe it fundamentally alters our results, particularly given our use of ANCOVAs which account for differences in initial bias when considering the different measures pertaining to perception of our robots.

In terms of B_2, although with a lower order of magnitude, there was a significant difference between the scores from the different participant gender on the pre-test measure ($H(1) = 4.517$, $p = 0.034$). Women agreed much more strongly with the need to change both gender trends, i.e. that it is important to encourage women to study STEM and to encourage men to get involved with childcare with scores of $M = 1.262$, $SD = 0.578$ and $M = 1.461$, $SD = 0.868$ respectively. No statistical significant result is revealed for B_2 post-hoc answers.

Finally, specifically considering any immediate change within participants induced by watching the video, no significant difference was found between participants pre- and post-test B_i measures across any of our manipulations, nor across participant gender. We were thus unable to replicate any evidence that observing our video stimulus might be enough, in of itself, to impact participant' bias [6].

Concerning the relationship between participant bias and their perception of the robot, Spearman's correlation analyses between each bias measure and the credibility and likeability ratings from participants are shown in Table 5. The most significant associations have been found between measure B_2 and post-hoc credibility and likeability. In all combinations, the correlations resulted in a rather low association, where the strongest correlation in credibility is between $B_{2,pre}$ and $\text{Credibility}_{post}$ ($\rho = -0.228$, $p < 0.001$) accounting for $\sim 5\%$ of the variance in post-hoc credibility; and between $B_{2,post}$ and $\text{Likeability}_{post}$ for likeability ($\rho = -0.159$, $p < 0.01$) accounting for $\sim 2.5\%$ of its variance. Notably (and perhaps unsurprisingly) all but the least significant coefficients identify a negative correlation between the participant bias score and their rating of the robot; i.e. the higher the participants' bias, the lower they rated the robot's credibility and likeability.

Table 5 Spearman's correlation analysis between bias measures B_i with ratings of credibility and likeability (pre- and post-hoc)

Spearman's correlation (ρ)	Credibility_{pre}	$\text{Credibility}_{post}$
(a)		
$B_{1,pre}$	0.039	−0.010
$B_{1,post}$	0.098	0.060
$B_{2,pre}$	−0.091	−0.228***
$B_{2,post}$	−0.088	−0.199***
Spearman's correlation (ρ)	Likeability_{pre}	$\text{Likeability}_{post}$
(b)		
$B_{1,pre}$	0.052	−0.011
$B_{1,post}$	0.115*	0.074
$B_{2,pre}$	−0.023	−0.133*
$B_{2,post}$	−0.007	−0.159**

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5 Discussion

Our analysis suggests complex interactions between robot gendering, interactant gender, and observer gender that can seem to impact on how a social robot is perceived when discussing (and challenging) gender stereotyping. Taking into account the factors such as who a robot is talking *to*, and *about* may help increase its credibility and its potential (perceived and actual) impact when attempting to challenge and dispel gender stereotypes.

5.1 Gender Doesn't Matter, Until it Does

Credibility and likeability ascribed to the robot after watching the initial introductory video did not vary across robot gendering nor participant gender. This aligns to recent work which failed to find evidence that robot gendering influences

e.g. perceptions of suitability for specific tasks [3]. This suggests that robot gendering does not seem to influence first impressions of a social robot, i.e. that there is no universal preference for male or female robot gendering. It might also be the case that this remains true in other social robot applications for which gender is less pertinent to the interaction.

However, contrasting with our initial pre-hoc results, our post-hoc results replicate Jackson et al.'s findings regarding complex interactions between robot gendering, interactant gender and observer gender on watching a discussion based scenario, including a robot-actor disagreement, in which (gendered) politeness norms might be pertinent to the interaction [19]. Watching the same robot actively challenge a stereotype had an impact on participants' ascription of credibility and likeability, in a way that suggests complex interactions across our gender manipulations. For example, women ascribed the male robot more credibility and likeability than men did when that robot was challenging the stereotype about men, and a similar effect was also seen when the male robot was challenging a male actor. Thus, generally, women appeared to be more impressed than men by our stereotype-challenging robot; and men seemed particularly less impressed with the robot when it was trying to challenge stereotypes *about* men or propagated *by* a man. This raises interesting questions about how, as Strengers and Kennedy call for [5], to design robots which challenge gender stereotypes without alienating (male) users.

The female presenting version of our robot was ascribed significantly less credibility when talking about women in STEM, specifically the same scenario investigated by [6]. This combination also appeared to result in the robot being ascribed the least potential real-world impact. Overall then, our results could be used to suggest that male robot gendering might be preferential/have more impact than female gendering in the context of challenging gender stereotypes through persuasive interactions, likely reflecting those gendered expectations which also generally result in men being considered more influential than women [31]. In short, gendered robots are judged differently by men and women based on who they are talking to and about. This raises an interesting question at the intersection of this work and Winkle et al.'s [6]: in order to have the best possible chance at tackling gender stereotyping in the wild, do we prioritise the use of robot gendering to maximise direct persuasiveness and impact? Or rather prioritise the use of robot gendering to demonstrate and normalise gender norm-breaking behaviour? Using a male-presenting robot might fall into either category, as it risks propagating norms about whose voice is listened to and has power; but could also be used to model alternative male behavioural norms.

Notably, unlike our other measures, our questions about the robot's perceived impact were only posed at the end of the study (this was done in order to avoid giving too much information to the participant before watching the video) so we cannot assert if the difference in perceived impact was already present from the beginning, i.e. representing a universal perception that the male presenting has more potential impact than female presenting robot. This would not align to our initial pre-hoc results on the (lack of) difference in credibility and likeability ascribed to the robot, nor with research done in psychology suggesting men, in general, are not perceived as more persuasive than women [56]. As such, we hypothesise that this difference specifically arises from our actual scenario and dialogue. We speculate that participants' attribution of greater impact to the male presenting robot might simply be reflecting the idea that men do have more impact in a patriarchal society, which, if so, again brings us to a question of whether we are better to 'lean in' to this, and utilise male robot gendering to the end of better effecting behaviour change, or rather follow [6] in explicitly using female presenting robots to challenge that status quo. Alternatively, as alluded to above, perhaps there is a way to actually leverage the perceived impact of male presenting robots by using them to challenge traditional stereotypes/model alternative forms of male behaviour. This seems a particularly interesting avenue for future work, especially given our findings regarding men being most put off by a robot challenging stereotyping about men and or/from a man, and is in-line with Strengers and Kennedy's calling for examination of how male-presenting robots engaging in traditionally feminised labour might represent a positive step towards queering of assistive technologies [5].

5.2 On the Potential for Impacting Observer Biases with HRI

Concerning *actually* challenging gender stereotypes, i.e. having an effect on the participants' pre-existing biases, neither of our robots appeared to induce any significant change in participant bias. Of course, this might not be surprising given that attitude and behaviour change is typically a longitudinal process [27], but this does mean we failed to replicate the findings in [6] where the robot succeeded in influencing some biases of the young teenagers watching the video. We hypothesise that adult populations are perhaps more ingrained in their biases, making it more difficult to change them with such a short intervention.

By also examining the correlation between participant gender bias and perceptions of our (stereotype-challenging) robot, we hoped to further probe the potential our robot might have for actually influencing biased individuals. Our hypothesis was that those with high gender bias would be less impressed by a robot trying to tackle those biases, in turn meaning that robots might have less chance of influencing them (given that perceptions of credibility and likeability are known to correlate with persuasiveness in human com-

municators [27]). Whilst we did evidence of exactly such a correlation, it was not as strong as might be assumed, although future work might look at more subtle interventions designed to tackle gender stereotyping without "turning off" those more biased individuals by raising the topic so directly.

6 Conclusion

With this work, we examined the influence of gender in HRI, specifically exploring interactions between robot gendering, interactant gender and observer gender on perceptions of a social robot when challenging gender stereotypes about men and women. While robot gendering did not have an effect on perceived persuasiveness after a first, short, introductory interaction, this changed once the robot spoke up in favour of challenging gender stereotypes, and provided a rationale-based counter-argument towards a (seemingly gender-biased) interactant. We do therefore find evidence that robot gender presentation can influence credibility, likeability, and likely therefore persuasiveness in the context of trying to change attitudes regarding gender bias. However, our findings suggest that biases, norms and tendencies from a patriarchal culture might be ingrained into participants' answers, pointing to a difficult intersection for designers that want to combat such stereotypes through technology, as we need to negotiate maximising persuasiveness and acceptance whilst avoiding (i) the propagation of harmful norms and (ii) missed opportunities for demonstrating and normalising gender norm-breaking behaviour.

6.1 Limitations and Future Work

Our work starts from an assumption of robot use, and as such is primarily concerned with the impact of different robot design choices. Future work might examine how a robot intervention compares and contrasts with e.g. a human and/or computer-based intervention in order to comment on the efficacy and specificities of HRI in the context of challenging stereotypes. On cultural specificity, it should also be noted that the stereotypes we investigate are taken to be typical of western society, per the background of the research team background and the demographics of our participants, such that our interaction scenario and experimental results should not be considered universally generalisable.

On further limitations, measures regarding potential impact might be revisited in pre-hoc testing in the future, as there might be a difference if that was asked before the interaction. To better comment on real-world impact potential, an in-person study would best examine whether our findings correlate with real attitude and/or behaviour change when the subject interacts with (or observes) the robot, although we note this would still likely require longitudinal testing. It would also be interesting to investigate further with a controlled more biased vs less biased population to have a better understanding of their perception of this kind of application of social robots (and whether we can really impact those who are most biased to begin with). Finally, more qualitative data collection and analysis would help to identify specific thought processes underlying participant impressions that could better ground and explain our findings, as well as provided much needed space for participatory discussion and reflection on these pressing issues of gender fairness in HRI which cannot be 'solved' by researcher-led quantitative experimental work alone.

Funding Open access funding provided by Royal Institute of Technology. This work was partially funded by grants from the Swedish Research Council (2017-05189), the Swedish Foundation for Strategic Research (SSF FFL18-0199), the S-FACTOR project from NordForsk, the Digital Futures research Center, the Vinnova Competence Center for Trustworthy Edge Computing Systems and Applications at KTH, and the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

Data availability The datasets generated during and/or analysed during the current study are available in the "The Right (Wo)Man for the Job? Exploring the Role of Gender when Challenging Stereotypes with a Social Robot"OSF repository, https://osf.io/uds84/?view_only=7d6ba5d7e87d4f1baa32edc4ef984823.

References

1. Ghazali AS, Ham J, Barakova EI et al (2018) Effects of robot facial characteristics and gender in persuasive human–robot interaction. Front Robot AI 5:73
2. Eyssel F, Hegel F (2012) (S)he's got the look: gender stereotyping of robots. J Appl Soc Psychol 42(9):2213–2230. https://doi.org/10.1111/j.1559-1816.2012.00937.x
3. Bryant D, Borenstein J, Howard A (2020) Why should we gender? The effect of robot gendering and occupational stereotypes on human trust and perceived competency. In: Proceedings of the 2020 ACM/IEEE international conference on human–robot interaction, pp 13–21
4. West M, Kraut R, Ei Chew H (2019) I'd blush if I could: closing gender divides in digital skills through education. UNESCO, Tech. rep
5. Strengers Y, Kennedy J (2020) The smart wife: why Siri, Alexa, and other smart home devices need a feminist reboot. MIT Press

6. Winkle K, Melsión GI, McMillan D, et al (2021) Boosting robot credibility and challenging gender norms in responding to abusive behaviour: A case for feminist robots. In: Companion of the 2021 ACM/IEEE international conference on human–robot interaction, pp 29–37
7. Ham J, Midden CJ (2014) A persuasive robot to stimulate energy conservation: the influence of positive and negative social feedback and task similarity on energy-consumption behavior. Int J Soc Robot 6(2):163–171
8. Maeda R, Brščić D, Kanda T (2021) Influencing moral behavior through mere observation of robot work: video-based survey on littering behavior. In: Proceedings of the 2021 ACM/IEEE international conference on human–robot interaction, pp 83–91
9. Sullivan A, Bers MU (2019) Investigating the use of robotics to increase girls' interest in engineering during early elementary school. Int J Technol Des Educ 29(5):1033–1051. https://doi.org/10.1007/s10798-018-9483-y
10. Siegel M, Breazeal C, Norton MI (2009) Persuasive robotics: the influence of robot gender on human behavior. In: 2009 IEEE/RSJ international conference on intelligent robots and systems, pp 2563–2568, https://doi.org/10.1109/IROS.2009.5354116
11. Thellman S, Hagman W, Jonsson E, et al (2018) He is not more persuasive than her: No gender biases toward robots giving speeches. In: Proceedings of the 18th international conference on intelligent virtual agents. Association for computing machinery, New York, IVA '18, pp 327–328, https://doi.org/10.1145/3267851.3267862
12. Rea DJ, Wang Y, Young JE (2015) Check your stereotypes at the door: an analysis of gender typecasts in social human-robot interaction. In: Tapus A, André E, Martin JC, et al (eds) Social robotics. Springer International Publishing, Cham, Lecture Notes in Computer Science, pp 554–563, https://doi.org/10.1007/978-3-319-25554-555
13. Reich-Stiebert N, Eyssel F (2017) (Ir)relevance of gender? On the influence of gender stereotypes on learning with a robot. In: Proceedings of the 2017 ACM/IEEE international conference on human–robot interaction. Association for Computing Machinery, New York, HRI '17, pp 166–176, https://doi.org/10.1145/2909824.3020242
14. Chita-Tegmark M, Lohani M, Scheutz M (2019) Gender effects in perceptions of robots and humans with varying emotional intelligence. In: 2019 14th ACM/IEEE international conference on human–robot interaction (HRI), pp 230–238, https://doi.org/10.1109/HRI.2019.8673222
15. Nass C, Moon Y, Green N (1997) Are machines gender neutral? Gender-stereotypic responses to computers with voices. J Appl Soc Psychol 27(10):864–876. https://doi.org/10.1111/j.1559-1816.1997.tb00275.x
16. Nomura T (2017) Robots and gender. Gend Genome 1(1):18–25
17. Crowell CR, Villano M, Scheutz M, et al (2009) Gendered voice and robot entities: perceptions and reactions of male and female subjects. In: 2009 IEEE/RSJ international conference on intelligent robots and systems, IEEE, pp 3735–3741
18. Paetzel M, Peters C, Nyström I, et al (2016) Congruency matters-how ambiguous gender cues increase a robot's uncanniness. In: International conference on social robotics, Springer, pp 402–412
19. Jackson RB, Williams T, Smith N (2020) Exploring the role of gender in perceptions of robotic noncompliance. In: Proceedings of the 2020 ACM/IEEE international conference on human–robot interaction. Association for Computing Machinery, New York, pp 559–567
20. Winkle K, Lemaignan S, Caleb-Solly P, et al (2019) Effective persuasion strategies for socially assistive robots. In: 2019 14th ACM/IEEE international conference on human–robot interaction (HRI), IEEE, pp 277–285
21. Chidambaram V, Chiang YH, Mutlu B (2012) Designing persuasive robots: how robots might persuade people using vocal and nonverbal cues. In: Proceedings of the seventh annual ACM/IEEE international conference on human–robot interaction. ACM, pp 293–300
22. Nakagawa K, Shiomi M, Shinozawa K, et al (2011) Effect of robot's active touch on people's motivation. In: Human–robot interaction (HRI), 2011 6th ACM/IEEE international conference on. IEEE, pp 465–472
23. Wills P, Baxter P, Kennedy J, et al (2016) Socially contingent humanoid robot head behaviour results in increased charity donations. In: 2016 11th ACM/IEEE international conference on human–robot interaction (HRI), IEEE, pp 533–534
24. Jackson RB, Williams T (2019) Language-capable robots may inadvertently weaken human moral norms. In: 2019 14th ACM/IEEE international conference on human–robot interaction (HRI), IEEE, pp 401–410
25. Rudman LA, Ashmore RD, Gary ML (2001) Unlearning automatic biases: the malleability of implicit prejudice and stereotypes. J Personal Soc Psychol 81(5):856
26. Dasgupta N, Asgari S (2004) Seeing is believing: exposure to counterstereotypic women leaders and its effect on the malleability of automatic gender stereotyping. J Expe Soc Psychol 40(5):642–658
27. Gass R, Seiter J (2018) Persuasion: social influence and compliance gaining. https://doi.org/10.4324/9781315209302
28. Cacioppo JT, Petty RE (1984) The elaboration likelihood model of persuasion. In: ACR north American advances
29. Petty RE, Briñol P (2011) The elaboration likelihood model. In: Handbook of theories of social psychology 1:224-45
30. Lockheed ME (1985) Sex and social influence: a meta-analysis guided by theory. In: Berger J, Zelditch M Jr (eds) Status, rewards, and influence: how expectations organize behavior, pp 406–429
31. Carli LL (2001) Gender and social influence. J Soc Issues 57(4):725–741
32. Ward DA, Seccombe K, Bendel R, Carter LF (1985) Cross-sex context as a factor in persuasibility sex differences. Soc Psychol Q 269–276
33. Cody MJ, Seiter JS, Montagne-Miller Y (1995) Men and women in the marketplace. In: Kalbfleisch P, Cody MJ (eds) Gender, power, and communication in human relationships. Lawrence Erlbaum Associates, Hillsdale, pp 305–329
34. Block K (2012) Communal male role models: how they influence identification with domestic roles and anticipation of future involvement with the family. In: University of British Columbia's undergraduate journal of psychology
35. Ellemers N (2018) Gender stereotypes. In: Annual review of psychology
36. Riggs JM (1997) Mandates for mothers and fathers: perceptions of breadwinners and care givers. In: Sex roles
37. Brescoll VL, Uhlmann EL (2005) Attitudes toward traditional and nontraditional parents. In: Psychology of women quarterly
38. Schein VE, Mueller R, Lituchy T, Liu J (1996) Think manager-think male: a global phenomenon? J Organ Behav 17(1):33–41
39. Lee Badgett MV, Folbre N (2003) Job gendering: occupational choice and the marriage market. Ind Relati: A J Econ Soc 42(2):270–298
40. Olsson M, Martiny SE (2018) Does exposure to counterstereotypical role models influence girls' and women's gender stereotypes and career choices? A review of social psychological research. Front Psychol 9:2264
41. Stout JG, Dasgupta N, Hunsinger M, McManus MA (2011) STEMing the tide: using ingroup experts to inoculate women's self-concept in science, technology, engineering, and mathematics (STEM). J Personal Soc Psychol 100(2):255
42. Curry AC, Robertson J, Rieser V (2020) Conversational assistants and gender stereotypes: public perceptions and desiderata for voice personas. In: Proceedings of the second workshop on gender bias in natural language processing, pp 72–78

43. Moradbakhti L, Schreibelmayr S, Mara M (2022) Do men have no need for"feminist"artificial intelligence? Agentic and gendered voice assistants in the light of basic psychological needs. Front Psychol 13
44. Tanqueray L, Paulsson T, Zhong M, Larsson S, Castellano G (Accepted/In press). Gender fairness in social robotics: exploring a future care of peripartum depression. In: Proceedings of the 2022 ACM/IEEE international conference on human–robot interaction: Alt.HRI—our robotics futures: a time capsule association for computing machinery (ACM)
45. Reich-Stiebert N, Eyssel F (2017) (ir) relevance of gender? on the influence of gender stereotypes on learning with a robot. In: 2017 12th ACM/IEEE international conference on human–robot interaction (HRI), IEEE, pp 166–176
46. Katie D (2021) Women in stem, percentages of women in stem statistics. https://www.stemwomen.co.uk/blog/2021/01/women-in-stem-percentages-of-women-in-stem-statistics. Accessed 09 June 2021
47. Díaz-García C, González-Moreno A, Jose Saez-Martinez F (2013) Gender diversity within r&d teams: its impact on radicalness of innovation. Innovation 15(2):149–160
48. The European Institute for Gender Equality (EIGE) (2019) How gender equality in stem education leads to economic growth. https://eige.europa.eu/gender-mainstreaming/policy-areas/economic-and-financial-affairs/economic-benefits-gender-equality/stem
49. Chronholm A (2007) Fathers'experience of shared parental leave in Sweden. Recherches Sociologiques et Anthropologiques 38(38–2):9–25
50. Rush M (2015) Between two worlds of father politics: USA or Sweden? Manchester University Press
51. Hentschel T, Heilman ME, Peus CV (2019) The multiple dimensions of gender stereotypes: a current look at men's and women's characterizations of others and themselves. Front Psychol 10:11
52. Chin H, Molefi LW, Yi MY (2020) Empathy is all you need: how a conversational agent should respond to verbal abuse. In: Proceedings of the 2020 CHI conference on human factors in computing systems, pp 1–13
53. Bartneck C, Kulić D, Croft E et al (2009) Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. Int J Soc Robot 1(1):71–81
54. Fiske ST, Cuddy AJ, Glick P. (2007) Universal dimensions of social cognition: Warmth and competence. In: Trends in cognitive sciences. pp 77–83
55. JASP Team (2021). JASP (Version 0.16)[Computer software]
56. Holtgraves T, Lasky B (1999) Linguistic power and persuasion. J Lang Soc Psychol 18(2):196–205

Alessio Galatolo is PhD student at the Uppsala Social Robotics Lab, Uppsala University (Sweden). He completed his master's degree in Machine Learning at KTH Royal Institute of Technology (Sweden). His research interests concern the use of AI and Machine Learning in Social Robotics with a particular interest on robots' social intelligence and how they can be personalised to a particular user. Alessio is also interested in feminist design for social robots and their use to fight common biases and inequalities in society.

Gaspar I. Melsíon was, at the time of writing, a PhD student at KTH Royal Institute of Technology (Sweden) where he also worked as a Research Engineer. His research project concerned Explainable AIs. He obtained his Master of Science in Engineering jointly from KTH and UPC - ETSETB (Spain). Gaspar is now working as a Software Engineer at Northvolt (Sweden).

Iolanda Leite is an Associate Professor at the School of Electrical Engineering and Computer Science at KTH Royal Institute of Technology. She holds a PhD in Information Systems and Computer Engineering from IST, University of Lisbon. Prior to joining KTH, she had postdoctoral appointments at Yale University and at Disney Research. Her research goal is to develop social robots that can perceive, learn from and respond appropriately to people in real-world situations, allowing for truly efficient and engaging long-term interactions with people. She was awarded one of the individual grants by the Swedish Foundation for Strategic Research's Future Research Leaders program (2020-2025). Leite is a member of the HRI Steering Committee and from 2018 to 2021 was the Co-editor in Chief of AI Matters, the newsletter of the ACM Special Interest Group in Artificial Intelligence (SIGAI).

Dr. Katie Winkle PhD is an Assistant Professor in Social Robotics at Uppsala University (Sweden). Previous to this she held a Digital Futures postdoctoral research fellowship at KTH Royal Institute of Technology (Sweden), after completing her PhD at the Bristol Robotics Laboratory (UK). Her research is concerned with delivering 'real-world' effective, ethical human-robot interactions; with recent research projects being focused on expert-led robot development of social robots in health and education settings alongside developing notions of Feminist human-robot interaction.