Models of (Often) Ambivalent Robot Stereotypes: Content, Structure, and Predictors of Robots' Age and Gender Stereotypes

Giulia Perugia
g.perugia@tue.nl
Dept. of Industrial Engineering & Innovation Sciences
Eindhoven University of Technology
Eindhoven, Netherlands

Latisha Boor
l.a.m.boor@tilburguniversity.edu
Tilburg School of Humanities & Digital Sciences
Tilburg University
Tilburg, Netherlands

Laura van der Bij
l.n.v.d.bij@student.tue.nl
Dept. of Industrial Engineering & Innovation Sciences
Eindhoven University of Technology
Eindhoven, Netherlands

Okke Rikmenspoel
o.m.rikmenspoel@student.tue.nl
Dept. of Industrial Engineering & Innovation Sciences
Eindhoven University of Technology
Eindhoven, Netherlands

Robin Foppen
r.a.m.foppen@student.tue.nl
Dept. of Industrial Engineering & Innovation Sciences
Eindhoven University of Technology
Eindhoven, Netherlands

Stefano Guidi
stefano.guidi@unisi.it
Dept. of Social, Political & Cognitive Sciences
University of Siena
Siena, Italy

ABSTRACT

This study focused on investigating the content, structure, and predictors of robots' stereotypes. We involved 120 participants in an online study and asked them to rate 80 robots on communion, agency, suitability for female and suitability for male tasks. In line with the stereotype content model, we discovered that robots' stereotypes are described by two dimensions, communion and agency, which combine to form univalent (e.g., low communion/low agency), as well as ambivalent clusters (e.g., low communion/high agency). Moreover, we found out that a robot's stereotypical appearance has a role in activating stereotypes. Indeed, in our study, female robots featuring appearance cues socio-culturally associated with femininity (e.g., eyelashes or apparel) were perceived as more communal, and juvenile robots featuring appearance cues tapping into the baby schema (e.g., cartoony eyes) were perceived as more communal, less agentic, and less suited to perform tasks. Given the renowned relationship between stereotyping, prejudice and discrimination, the causal link between appearance and stereotyping we establish in this paper can help HRI researchers disentangle the relation between robots' design and people's behavioral tendencies towards them, including proneness to harm.

CCS CONCEPTS

• **Human-centered computing** → HCI theory, concepts and models; **Empirical studies in HCI**.

KEYWORDS

HRI, Social Robots, Stereotype Content Model, Bias, Gender

HRI '23, March 13–16, 2023, Stockholm, Sweden

ACM ISBN 978-1-4503-9964-7/23/03.
https://doi.org/10.1145/3568162.3576981

ACM Reference Format:
Giulia Perugia, Latisha Boor, Laura van der Bij, Okke Rikmenspoel, Robin Foppen, and Stefano Guidi. 2023. Models of (Often) Ambivalent Robot Stereotypes: Content, Structure, and Predictors of Robots' Age and Gender Stereotypes. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction (HRI '23), March 13–16, 2023, Stockholm, Sweden.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3568162.3576981

1 INTRODUCTION

Social categorization, the attribution of social categories such as age and gender to an agent, can elicit stereotyping [13], oversimplified views of a social group [23], and stereotyping can often result in prejudice and bias [11, 23]. In the field of Human-Robot Interaction (HRI), there is no conclusive evidence as to whether people actually stereotype robots based on perceived age and gender. More often than not, male robots are perceived as more agentic than female robots [16, 39], and female robots are perceived as more communal than male ones [5, 16, 27]. There are, however, a number of instances where differences in stereotypical traits fail to emerge [4, 33, 42, 44], and little is known about the stereotypes elicited by a robot's perceived age. Indeed, in spite of robots being known to differ in terms of perceived age [37], HRI research has only rarely touched upon the effects that this can have on people's perceptions of a robot, especially in relation to stereotyping. We suspect that the lack of conclusive results on robots' stereotypes could be due to the only seldom focus on the potentially ambivalent nature of stereotypes and to the habit of using only one robotic platform per study [27].

With the study reported in this paper, we aim to: (1) Identify the (often ambivalent) content and structure of robot stereotypes (both in terms of stereotypical traits and stereotypical tasks); (2) Determine whether the appearance attributes of a robot could have a role in influencing stereotyping, and whether this role is mediated by the robot's perceived gender and age; and (3) Disclose what individual characteristics make us humans more prone to stereotype robots (e.g., hostile and benevolent sexism). With respect to Reeves et al. [43], who carried out a similar study, we focus on a more restricted number of robots with less heterogeneous appearances (i.e., the robots in this study are mostly humanoid, while those in Reeves et al. were also animal- or object-like). Besides, we ask more people (30 vs. 10 in [43]) to rate the same robot in terms of stereotypes so as to reach more stable ratings [37, 40]. Last, we include task suitability as a relevant stereotype variable, and attempt to understand which individual characteristics affect stereotyping.

The research featured in this paper could contribute to unravel the functioning of stereotyping in HRI, especially in relation to the robot's design. Given the demonstrated relationship between stereotypes, prejudice, and discrimination [11, 18], reaching a systematic understanding of how people attribute stereotypes to robots, and on the basis on which appearance features, could help us predict the ways people will behave towards robots based on their design characteristics, including whether people would be more or less prone to bully or mistreat them [32]. The data collected in this study are made publicly available in the dataset attached to this paper.

1.1 Stereotypical Traits

In this paper, we adopt the Stereotype Content Model (SCM) [23] as a theoretical framework. The SCM is a model describing the content and structure of stereotypes. According to it, when we encounter a person for the first time, the first characteristic we appraise is their *warmth*, we want to know whether that person has good or bad intentions [23]. Only later we focus on their *competence*, aiming to evaluate whether that person actually has the ability to enact those intentions [23]. According to Fiske et al., warmth and competence are the characteristics that inform our perceptions of others [23], and the material stereotypes are made of.

The SCM posits that warmth and competence often operate ambivalently. The two dimensions can indeed form *univalent compounds*, either completely positive (high warmth/high competence) or completely negative (low warmth/low competence), but can also form *ambivalent compounds* with warmth overpowering competence, or vice versa [23]. In a series of studies [19, 20, 22, 23], and further replications of them, Fiske et al. have demonstrated that most *human* social groups fit in the four quadrants of a plot described by the two dimensions of warmth and competence. The cultural defaults in society (e.g., middle-class people) are the only ones eliciting univocally positive perceptions on both dimensions, whereas those living at the margins of society, such as homeless people, are the only ones eliciting univocally negative perceptions [23]. All the other groups are perceived ambivalently. Feminist women, for instance, elicit perceptions of high competence and low warmth; housewives and elderly people perceptions of low competence and high warmth [23]. In this paper, we follow the SCM structure but focus on communion and agency [2] as stereotype dimensions. This is to specifically keep continuity with previous HRI research. Communion and warmth, and agency and competence, are by large the same constructs. Agency, similar to competence, has to do with capability and goal-orientedness [1, 2]; communion, similar to warmth, with social desirability and friendliness [1, 2].

The study of stereotypes, especially ambivalent ones, is fundamental to gain a thorough understanding of the social dynamics occurring between agents. Indeed, communion and agency perceptions have been linked to specific affective reactions (or prejudices) an individual can attract [23]. While those high in both communion and agency elicit *admiration and pride*, *contempt and disgust* are reserved to those that are neither communal nor agentic (contemptuous prejudice), *pity and sympathy* target those that are communal but not agentic (paternalistic prejudice), and *envy and jealousy* those that are agentic but not communal (envious prejudice) [23]. Besides, the Behaviors from Intergroup Affect and Stereotypes (BIAS) Map [11] specifically linked the stereotype content, and the emotions it elicits (prejudice), with specific discriminatory tendencies people enact towards others: *passive facilitation* (e.g., convenient cooperation), *active facilitation* (e.g., help), *passive harm* (e.g., withdrawal of support), and *active harm* (e.g., verbal harassment). In HRI, a systematic understanding of the content of the stereotypes social robots elicit could help us predict the affective reactions they might attract and the mistreatment they might receive, including bullying. In this paper, we focus on the following research questions (**RQ1**):

RQ1a *What is the content and structure of robot's stereotypes (in terms of communion and agency)?*

RQ1b *To what extent does the robot's perceived gender affect the activation of stereotypes (in terms of communion and agency)?*

RQ1c *To what extent does the robot's perceived age affect the activation of stereotypes (in terms of communion and agency)?*

1.2 Stereotypical Tasks

Age and gender stereotypes do not only prescribe which traits certain social groups should have (i.e., communion and agency), but also which tasks these groups are suitable for [14]: men working outside the home, women within the home [24]; women being mothers, teachers, and nurses [14], men carrying out instrumental behaviors related to task accomplishment and resource acquisition [14]. Since people are often sanctioned for behaviors that are not in line with their gender and age roles [9], they often abide by the roles prescribed by society [14]. These role expectations might be transferred to robots [7], with robot's perceived gender affecting the types of tasks a robot is deemed suitable for, while robot's perceived age determining whether a robot is deemed suitable to carry out tasks at all. A literature review analyzing the influence of gendering practices on people's perceptions and interactions with humanoid robots has found evidence for this hypothesis at least in relation to the robot's perceived gender [38], with female robots, similar to women, being considered more suited for tasks, such as childcare, household maintenance, and elderly care [4, 5, 16, 36], and male robots, similar to men, being considered more suited for tasks, such as guarding the house, transporting goods, and handcrafting [4, 5, 16, 36]. Following these results, in this paper, we also focus on a second set of research questions (**RQ2**):

RQ2a *What is the content and structure of robot's stereotypes (in terms of task suitability)?*

RQ2b *To what extent does the robot's perceived gender affect the activation of stereotypes (in terms of task suitability)?*

RQ2c *To what extent does the robot's perceived age affect the activation of stereotypes (in terms of task suitability)?*

1.3 The Role of Appearance

Beyond the mere social categories attributed to robots (i.e., gender and age), the very appearance attributes used to suggest these social categories might have a role in eliciting stereotypes [3, 28, 37]. Roboticists use multiple cues to imbue gender and age into robots,

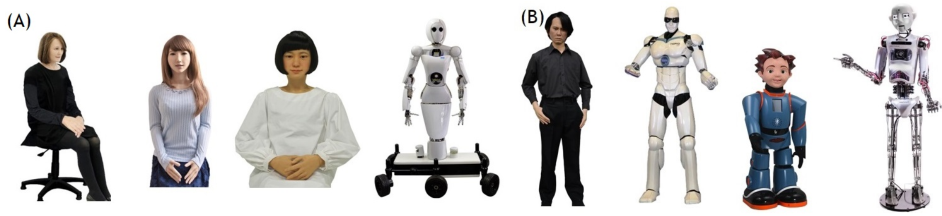

Figure 1: Female (A) and male robots (B) Left to right: Nadine, Erica, Kodomoroid, Aila, Geminoid, Topio, Zeno, Robothespian.

sometimes relying on extremely stereotyped appearances [3, 38, 39, 45] that mirror the 'normativized body' (the norm of bodily appearance for a certain gender and age presented by the media in a certain culture) [36, 41]. If we take a look at the pictures of the Anthropomorphic roBOTs (ABOT) database (see fig. 1), female robots are usually presented as seated, the legs closed, the arms and hands on the lap, a slender, interiorized body (hidden mechanical parts [45]), long hair and feminine apparel (see fig. 1A), while male robots are presented as standing, the legs apart, the arms and hands on the sides of the body, a sturdy muscular torso and exteriorized body (visible mechanical parts [45]), and short to no hair (see fig. 1B). When we focus on childlike and juvenile robots, instead, we cannot help but notice that their appearance features tap on the baby schema (a set of infantile features [25, 29, 30]) [37], with designers increasing the size of facial features, changing the proportions between body parts, and almost 'cartoonizing' the appearance of robots [34]. This neoteny of social robots might also be connected with the notion of *kawaii*, which is particularly widespread in Japan where many robots are produced [6, 31, 34, 35].

It almost seems like stereotypes about gender and age-related appearance become crystallized into humanoid robots' designs, and this might have a role in activating gender and age stereotypes, with robots tapping into gender and age-stereotypical appearances being more successful than others in eliciting the relevant stereotypes, due to their ability to activate stereotypical knowledge structures about age and gender in humans [15, 17]. To investigate this, we add the following research questions to RQ1a-c and RQ2a-c:

RQ1d *To what extent does the robot's appearance affect the activation of stereotypes (in terms of communion and agency)?*

RQ2d *To what extent does the robot's appearance affect the activation of stereotypes (in terms of task suitability)?*

Finally, it is not just the robot's appearance but also people's individual characteristics that could affect stereotyping [36]. The way we gender robots suggests that sexism and bias towards women could transfer to robots. Men perceive robots as more feminine and younger than women [37], and have the tendency to consider male roles as more specifically masculine than women [36]. Besides, hostile sexism (resenting women who violate gender norms [21]) and benevolent sexism (protecting women who adhere to traditional roles [21]) might be at play when people attribute stereotypical traits to robots, as well as appraise them as suitable for certain tasks [24]. To tackle the impact of participants' characteristics on stereotyping, we thus pose two last research questions:

RQ1e *How do the individual characteristics of the participants influence stereotyping (in terms of communion and agency)?*

RQ2e *How do the individual characteristics of the participants influence stereotyping (in terms of task suitability)?*

2 METHODS AND MATERIAL

2.1 Study Design

To investigate the attribution of stereotypical traits and tasks to humanoid robots, we carried out an online study inspired by Fiske et al. [23] and Eyssel and Hegel [16]. We selected 80 images of humanoid robots from the ABOT database [40], and randomly divided them into 4 groups of 20 images. We randomly allocated 30 participants to each of these groups, and asked them to rate the robots in the images on communion, agency, suitability for female, and suitability for male tasks. The order of presentation of the images was randomized across participants, while the order of presentation of stereotypical traits and tasks was randomized across robots. The images were kept in their original size.

2.2 Robot Selection

The 80 robots included in this study were sub-sampled from the 251 robots in the ABOT database following a two-step selection pipeline based on the following scores from the ABOT dataset [40]: the *humanlikeness* score (1 = not humanlike at all to 100 = just like a human) and the *surface look* (head hair, skin, nose, eyebrow, eyelashes, apparel, gender), *body manipulators* (torso, legs, arms, hands, fingers), and *facial features* (head, face, eyes, mouth) scores (0 = all features absent to 1 = all features present).

To get to the final sample of 80 robots, we first excluded the robots with extremely low or extremely high humanlikeness scores (< 20 and > 80; $N = 179$ robots left). Then, we selected the robots that had the 18 highest and 18 lowest surface look, body manipulators, and facial features scores ($N = 80$ robots left). The first criterion was aimed at excluding robots extremely similar to everyday objects or to humans, the second at ensuring a reasonable variety in the robots' appearance. Based on the ratings in the humanoid ROBOts - Gender and Age Perception (ROBO-GAP) dataset [37], 40 of the 80 robots included in the study were predominantly masculine (50%), 15 predominantly feminine (19%), and 25 predominantly gender neutral (31%); 25 had a mean age score higher than 30 (31%), 41 a mean age score comprised between 30 and 50 (51%), and 14 a mean age score comprised between 50 and 75 (18%). The list of robots can be found in the Supplementary Materials.

2.3 Participants

We recruited 120 participants via Prolific ($M_{age} = 26.76$, $SD_{age}= 9.04$, $Min_{age}= 18$, $Max_{age}= 74$; 58 women, 56 men, 5 non-binary

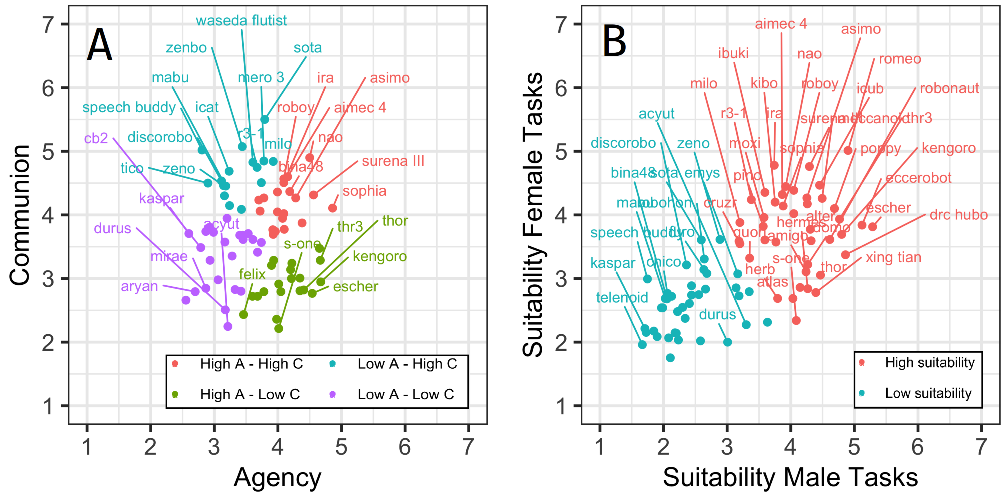

Figure 2: Results of the cluster analyses. Plots of the robots in the (A) Agency-Communion, and (B) Female-Male Task Suitability spaces colored according to the assigned cluster in the analysis.

people, 1 undisclosed, for more demographic details, see Appendix 1). All participants gave their informed consent to participate in the study and were compensated with £ 7.48. The study was approved by the Ethical Review Board of the Human-Technology Interaction group of Eindhoven University of Technology.

2.4 Measures and Procedure

Upon starting the questionnaire, participants were explained the general purpose of the study and were asked to provide their consent to participate. Once the consent was given, participants started rating the images of the robots. For each robot, they were shown a picture and were asked to rate their perceptions of the robot's: (1) *Communion* [10, 39], 5 items evaluated on a 7-point Likert scale (1 = strongly disagree to 7 = strongly agree; α = .95: The robot is affectionate, compassionate, tender, gentle, sympathetic); (2) *Agency* [10, 39], 5 items evaluated on a 7-point Likert scale (1 = strongly disagree to 7 = strongly agree; α = .89; The robot is able to defend its own beliefs, able to make decisions easily, willing to take a stand, has leadership abilities, has a strong personality); (3) *Suitability for female tasks* [8, 16], 5 items evaluated on a 7-point Likert scale (1 = not suited at all to 7 = extremely suited; α = .84; The robot is suited for use in caring for a child, household maintenance, preparing meals, providing therapy, taking care of the elderly); and (4) *Suitability for male tasks* [8, 16], 5 items evaluated on a 7-point Likert scale (1 = not suited at all to 7 =extremely suited; α = .76: The robot is suited for use in navigating a route, repairing a bike, steering a car, performing surgery, guarding my home).

When all 20 robots were evaluated, participants were asked to indicate their *age* (in numbers), *gender* (choosing from man, woman, non-binary, prefer not to say, and prefer to specify, followed by a blank field [47]), and *nationality* (open question), and indicate their *degree of familiarity with AI, robots, and science fiction* (7-point Likert scale, 1 = not familiar at all to 7 = very familiar) [37]. As a last step, we measured participant's *tendency to anthropomorphize* (3 items from Waytz et al.'s Individual Differences Anthropomorphism Questionnaire [49], e.g., To what extent does the average robot have consciousness? 10-point Likert scale from 1 = not at all to 10 = very much so, α = .75), *benevolent and hostile sexism* (12 items of the Ambivalent Sexism Inventory [46], e.g., "Many women have a quality of purity that few men possess", 7-point Likert scale, from 1 = strongly disagree to 7 = strongly agree; α_{BS} = .82, α_{HS} = .87), and *proneness to social desirability* (5 items from the Social Desirability Scale [48], e.g., "When I have made a promise, I keep it – no ifs, ands or buts", true/false answer; α = .41). This latter scale achieved low reliability, thus it was excluded from the analyses. Participants took $\approx$ 40 minutes to fill out the survey.

In our analyses, we also used the humanlikeness and appearance features scores from the ABOT database [40], which rate the humanlikeness of a robot, and the presence of body manipulators, surface look, and facial features (see section 2.2). Furthermore, we used the gender and age ratings from the ROBO-GAP dataset [37], which rate the extent to which people perceive the robots in the ABOT dataset as feminine, masculine, or gender neutral (7-point Likert scale), as well as their perceived age (from 1 to 100).

3 RESULTS

3.1 Data Screening

Before our actual analyses, we checked whether the participants in the four groups differed in their individual characteristics by performing six one-way ANOVAs. The results did not show any significant difference between groups in terms of familiarity with AI ($F(1, 116) = 0.69$, $p = .56$), robots ($F(1, 116) = 0.64$, $p = .59$), and sci-fi ($F(1, 116) = 0.52$, $p = .67$), nor in terms of tendency to anthropomorphize ($F(1, 116) = 0.34$, $p = .79$), benevolent ($F(1, 116) = 0.79$, $p = .50$), and hostile sexism ($F(1, 116) = 1.18$, $p = .32$). We used the mean ratings of the group of raters on communion, agency, stereotypical female and male tasks (for each robot) to perform the cluster and path analyses in 3.2 and 3.3. We resorted to individual ratings only to perform the regression analyses in 3.4.

3.2 The Content and Structure of Robot Stereotypes (RQ1a and RQ2a)

To disclose the content and structure of robot's stereotypes in terms of communion and agency (**RQ1a**) and task suitability (**RQ2a**), we performed two *k-means* cluster analyses. In the first analysis (see section 3.2.1), we used the aggregated ratings of agency and communion for clustering. In the second analysis, (see section 3.2.2), we used the aggregated ratings of suitability for stereotypical female and male tasks. The number of clusters was determined by plotting the Within-clusters Sum of Square (WSS) as a function of the number of clusters, and identifying the bend in the curve (similar to a scree plot, see WSS plots in Appendix 1).

3.2.1 Communion and Agency (RQ1a). For communion and agency, the last large change in the curve came in the bend between three and four clusters. We thus adopted a four-cluster solution. As visible in figure 2A, the four clusters corresponded to the clusters of the SCM: (1) **Cluster 1** included 24 robots with *low agency* and *low communion* ratings, among which Kaspar, Telenoid, and Flobi; (2) **Cluster 2** included 16 robots with *low agency* and *high communion* ratings, among which Zeno, Otto, and Robohon; (3) **Cluster 3** included 20 robots with *high agency* and *low communion* ratings, among which Kengoro, Socibot mini, and Robonaut; and (4) **Cluster 4** included 20 robots with *high agency* and *high communion* ratings, among which Asimo, Nao, and iCub. The list of robots in each cluster can be found in Appendix 1.

3.2.2 Suitability for Male and Female Tasks (RQ2a). For suitability for male and female tasks, the last large change in the curve came in the bend between one and two clusters. We thus adopted a two-cluster solution. As visible in figure 2B: (1) **Cluster 1** included 39 robots with *low female task suitability* and *low male task suitability*, among which Emys, Seer, Bina48, and Kismet; (2) **Cluster 2** included 41 robots with *high female task suitability* and *high male task suitability*, among which Sophia, Atlas, Robothespian, and Asimo. As shown in figure 2B, female and male task suitability were positively correlated. The two clusters thus represented high vs. low task suitability for both types of tasks. The list of robots in each cluster can be found in Appendix 1.

Table 1: *M* and *SD* for each of the identified clusters.

Cluster	**Agency** *M* (*SD*)		**Communion** *M* (*SD*)
1. Low A - Low C	3.16_a (0.34)[1]	=	3.32_a (0.49)
2. Low A - High C	3.39_a (0.34)	<	4.66_b (0.36)
3. High A - Low C	4.13_b (0.36)	>	2.90_c (0.33)
4. High A - High C	4.12_b (0.29)	=	4.17_d (0.33)
Clusters	**Suitability MT** *M* (*SD*)		**Suitability FT** *M* (*SD*)
1. Low suitability	2.41_a (0.50)	<=	2.59_a (0.46)
2. High suitability	4.12_b (0.53)	>	3.77_b (0.63)

[1]The symbols < or > indicate if the means significantly differ ($p < .05$) on agency and communion within groups; <= and >= indicate marginal significance ($p < .1$). Within each column, the means that do not share the same subscript significantly differ ($p < .05$). See Appendix 1 for the extended tables.

3.2.3 Validation of Cluster Solutions. To validate the cluster solutions, we performed two MANOVAs to compare (i) the scores of agency and communion, and (ii) suitability for female tasks and male tasks across clusters. The first MANOVA included the clusters identified in the first cluster analysis as fixed effects (4 levels). The second MANOVA included the clusters obtained from the second cluster analysis as fixed effects (2 levels). Both the 4-cluster solution for communion and agency and the 2-cluster solution for task suitability were validated. As follows, we report the results (the tables reporting the results of all the post-hoc analyses are featured in Appendix 1).

The first MANOVA showed a significant effect of cluster on communion and agency ($F(6, 152) = 61.0$, $p < .0001$). We thus ran follow-up univariate ANOVAs to explore which dependent variables significantly differed across clusters. The results revealed a significant effect of cluster on both agency ($F(3, 76) = 47.8$, $p < .0001$) and communion ($F(3, 76) = 79.95$, $p < .0001$). Independent-samples t-tests were then conducted to understand which group differed from which on the target dependent variables (p adjusted using Tukey's method). The results disclosed that, in the high agency clusters, robots were perceived as possessing significantly more agency than in the low agency clusters (all ps < .0001, see Table 1), while no significant differences in agency were found between pairs of clusters having the same level of agency but different levels of communion (both low agency: $p = .17$; both high agency: $p = .999$, see Table 1). The results for communion showed that in the high communion clusters robots were perceived as possessing significantly more communion than in the low communion clusters (all ps < .0001, see Table 1). However, they also showed significant differences in communion between clusters of robots with the same levels of communion but different levels of agency (both low communion: $p = .003$; both high communion: $p = .002$), with significantly lower communion ratings in the clusters with high agency (cluster 3 and 4) than in those with low agency (cluster 1 and 2). To identify differences between agency and communion *within* clusters, and thus verify the distinction between univalent and ambivalent stereotypes, we ran a series of paired-samples t-tests. In the two ambivalent clusters, agency and communion significantly differed from each other (all ps < .0001, see Table 1). In the two univalent clusters, they did not (high agency/high communion: $p = .54$; low agency/low communion: $p = .17$, see Table 1).

The second MANOVA showed a significant effect of cluster on suitability for female and male tasks ($F(2, 77) = 135.65$, $p < .0001$). Hence, we performed follow-up univariate ANOVAs to explore which dependent variable significantly differed across clusters. The results showed a significant effect of cluster on both suitability for stereotypical female ($F(1, 78) = 90.74$, $p < .0001$) and stereotypical male tasks ($F(1, 78) = 218.45$, $p < .0001$). The robots in the high suitability cluster were perceived as significantly more suited for both female and male tasks than the robots in the low suitability cluster (see table 1). To check whether differences in suitability for female and male tasks were present *within* the high suitability and low suitability clusters, we performed two paired-samples t-tests. We found that suitability for female tasks was marginally significantly higher than suitability for male tasks in the low suitability cluster ($p = .06$, see table 1), and significantly lower in the high suitability one ($p < .005$, see table 1).

3.3 Effect of Perceived Gender, Age, and Appearance on Stereotypes

To understand the extent to which robot's perceived gender, perceived age, and appearance influenced communion and agency (**RQ1b-d**), and task suitability (**RQ2b-d**), we tested six path models (PM), three using *communion and agency* as dependent variables (**PM1-PM3**), and three using *task suitability* as dependent variable (**PM4-PM6**). To account for the results of the cluster analysis, communion, agency, and task suitability were coded as binary variables in the models (low vs. high communion/agency/task suitability).

PM1 featured humanlikeness (HL), body manipulators (BM), surface look feature (SL), and facial features (FF) scores (from [40]), and robot's femininity and masculinity ratings (from [37]), as predictors, and *agency and communion* as dependent variables (see Fig. 3 for the structure of the model). **PM2** featured the same set of variables as PM1 but substituted robot's femininity and masculinity ratings with gender neutrality ratings (from [37]). **PM3** featured the same variables as PM1 and PM2, but included the robot's perceived age scores (from [37]) in lieu of robot's perceived gender ratings. **PM4**, **PM5**, and **PM6** replicated the structure of PM1, PM2, and PM3 respectively, but featured task suitability as dependent variable instead of communion and agency. Given the amount and complexity of the models, here we report only the significant results and avoid repeating the effects that do not vary across PM. The complete list of path models, fitness measures, and direct and indirect effects is available in the Appendix 2.

3.3.1 Communion and Agency (RQ1b-d). **PM1** showed an acceptable fit to the data ($X^2(2) = 8.006$, $p = .018$; $GFI = .983$, $SRMR = .082$). The presence of BM ($\beta = 0.52$, $p < .001$) and SL ($\beta = 0.67$, $p < .001$) significantly positively affected the robot's HL, which in turn significantly positively affected the robot's agency ($\beta = 0.48$, $p < .05$). This chain of effects entailed that BM and SL had both a significant positive indirect effect on agency via the robot's HL ($\beta_{BM} = 0.25$, $p < .05$, $\beta_{SL} = 0.32$, $p < .05$), with the effect of BM on agency being fully mediated by its effect on HL. BM also had a significant positive effect on masculinity ($\beta = 0.49$, $p < .001$) and a significant negative effect on femininity ($\beta = -0.33$, $p < .01$). Masculinity and femininity themselves had a significant effect on communion, negative the former, positive the latter ($\beta_M = -0.52$, $p < .001$; $\beta_F = 0.51$, $p < .001$). This resulted in a significant negative indirect effect of BM on communion via the robot's masculinity

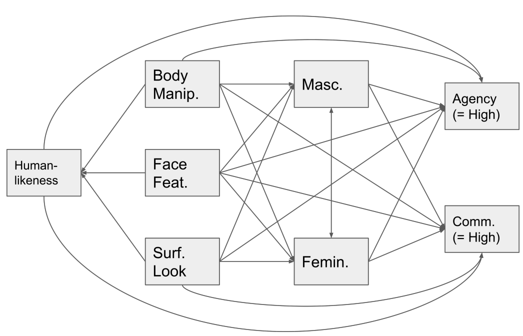

Figure 3: The first path model (PM1) tested in the study.

($\beta = -0.25$, $p < .001$), and via its femininity ($\beta = -0.17$, $p < .01$). The effect of SL on femininity was marginally significant ($\beta = 0.26$, $p = .06$), and so was the indirect effect of SL on Communion via femininity ($\beta = 0.13$, $p = .08$). Finally, FF had a significant effect on communion (positive, $\beta = 0.68$, $p < .01$) and agency (negative, $\beta = -0.36, p = .06$), but these effects were not mediated by the robot's perceived gender nor by its HL.

We then tested **PM2**. The fit of the model was good ($X^2(1) = 3.434$, $p = .064$; $GFI = .995$, $SRMR = .054$). Both BM and SL had a significant negative effect on perceived gender neutrality ($\beta_{BM} = -0.24$, $p < .001$, $\beta_{SL} = -0.81$, $p < .001$), yet perceived gender neutrality did not have any significant effect on agency and communion, nor did it act as a mediator of the effects of robot's appearance on the outcome variables. Last, we fitted **PM3**. The model had an excellent fit to the data ($X^2(1) = .091$, $p = .762$; $GFI = 1$, $SRMR = .010$). FF had a significant negative effect on perceived age ($\beta = -0.39$, $p < .05$), while SL a significant positive effect ($\beta = 0.32$, $p < .05$). Robot's perceived age had moderate to strong effects on both agency (positive, $\beta = 0.41$, $p < .001$) and communion (negative, $\beta = -0.90$, $p < .001$). These strong direct effects resulted in both FF and SL having significant indirect effects on agency ($\beta_{FF} = -0.16$, $p = .07$, $\beta_{SL} = 0.13$, $p < .05$) and communion via age ($\beta_{FF} = 0.35$, $p < .05$, $\beta_{SL} = -0.29$, $p < .05$).

3.3.2 Suitability for Tasks (RQ2b-d). As for task suitability, **PM4** showed an acceptable fit to the data according to some fit measures, and a poor fit according to other ($X^2(3) = 8.006$, $p = .018$; $GFI = .972$, $SRMR = .101$). BM had a significant positive effect on task suitability ($\beta = 0.82$, $p < .01$) and FF a marginally significant negative effect ($\beta = -0.29$, $p = .09$). All other predictors did not yield any significant effects on task suitability. The fit of **PM5** was good ($X^2(1) = 3.434$, $p = .064$; $GFI = .994$, $SRMR = .070$). However, similar to PM2, also for PM5, gender neutrality was not a significant predictor of the outcome variable, task suitability, and therefore not a possible mediator of the effects of appearance features on it. Last, we fitted **PM6**. The model had an excellent fit to the data ($X^2(1) = .091$, $p = .762$; $GFI = 1$, $SRMR = .013$). Age had a significant positive effect on task suitability ($\beta = 0.34$, $p < .001$), and fully mediated the negative effect of FF on task suitability ($\beta = -0.13$, $p = .08$). We also found a marginally significant indirect effect of SL on task suitability via age ($\beta = 0.11$, $p = .06$).

3.4 Effects of individual features

To understand how participants' individual characteristics influenced communion and agency (**RQ1e**) and task suitability (**RQ2e**), we conducted four linear mixed-effects hierarchical regressions using, respectively, agency (A), communion (C), suitability for female (SFT), and suitability for male tasks (SMT) as dependent variables, and participants' age, gender, tendency to anthropomorphize, familiarity with technologies, and hostile and benevolent sexism as predictors. For the models using suitability for female and male tasks as dependent variables, we also included agency and communion as predictors. All models included random adjustments of the intercepts by-robot and by-participant. The complete list of results, including the non-significant ones, is available in Appendix 2.

The results showed that participants' age had a significant positive effect on communion, agency, and suitability for male tasks

(B_A = 0.02, p = .029; B_C = 0.02, p = .030; B_{SMT} = 0.02, p = .017), while participants' tendency to anthropomorphize yielded significant effects on communion and agency (B_A = 0.11, p = .010; B_C = 0.10, p = .028). Moreover, they disclosed that participant's gender (i.e., male gender) had a significant negative effect on suitability for male tasks (B = −0.32, p = .021), hostile sexism a marginally significant negative effect on agency (B = −0.17, p = .053), and benevolent sexism a significant positive effect on agency (B = 0.19, p = .023) and a marginally significant one on male task suitability (B = 0.10, p = .062). When analyzing the effects of stereotypical traits on task suitability, we discovered that agency significantly positively predicted suitability for female and suitability for male tasks (B_{SFT} = 0.16, p < .001; B_{SMT} = 0.29, p < .001), while communion only suitability for female tasks (B = 0.40, p < .001).

4 DISCUSSION

4.1 Stereotypical Traits

4.1.1 Communion and Agency: Stereotype Content and Structure (RQ1a). The results of the first cluster analysis disclosed that **the content and structure of robot's stereotypes mirror the content and structure of human stereotypes** (**RQ1a**). We identified two *univalent compounds* of communion and agency - cluster 1 (low agency/low communion) and cluster 4 (high agency/high communion) - and two *ambivalent compounds* - cluster 2 (low agency/high communion) and cluster 3 (high agency/low communion).

If we take a look at each cluster (see Appendix 1), what we notice is that the robots in cluster 1 (*low agency/low communion*) have very disparate characteristics but all share a certain dull humanlikeness. Some of them have stylized humanlike features (e.g., Kaspar, Telenoid). Others have misplaced or missing body parts (e.g., the "head" on the legs in Walker and Cruzr). Still others are headless mechanical robots (e.g., Durus) or cartoony childlike ones (e.g., Adata, JD Human). If we interpret this result through the lens of the SCM [23] and BIAS Map [11, 32], we can postulate that these robots elicit affective reactions such as *contempt and disgust*, and discriminatory tendencies such as *passive* (e.g., excluding) *and active harm.*

Cluster 2 (*low agency/high communion*) features a less diverse set of robots than cluster 1, which seem to share two main characteristics: the robots in it are mostly juvenile (M_{age} = 26.5), and have an expressive face. Cluster 3 (*high agency/low communion*) features a quite homogeneous group of robots, too. However, as opposed to cluster 2, the robots in this cluster are all quite adultlike (M_{age} = 46.7), and have very few facial features. These two clusters represent ambivalent stereotypes. Based on Fiske et al. [22, 23] and Cuddy et al. [11], we can hypothesize that the robots in these clusters can elicit ambivalent emotions: *pity and sympathy* for cluster 2, and *envy and jealousy* for cluster 3 - as well as ambivalent discriminatory tendencies - *passive harm and active facilitation* for cluster 2, and *passive facilitation and active harm* for cluster 3. This is interesting, especially if we consider that the robots in these clusters are mostly differentiated by their perceived age and expressivity, with cluster 2 robots fitting the stereotypical "nice" robot of Disney movies (e.g., Wall-e, Eve), and cluster 3 the stereotypical "bad robot" of Hollywood movies (e.g., RoboCop, Terminator).

Finally, cluster 4 (*high agency/high communion*) features highly humanlike robots and androids. Similar to cluster 3, these robots are mostly full-bodied, and, similar to cluster 2, they almost all have a face. Interestingly though, while the robots in cluster 3 have an exteriorized body, the robots in cluster 4 have an interiorized body. Based on the SCM [23, 32] and BIAS Map [11], we can assume that these robots are perceived as ingroups, the cultural norm of social robotics. Hence, they elicit univalently positive emotions of *pride and admiration*, and do not attract discrimination tendencies, but *passive and active facilitation.* This result is interesting and bears resemblances with the uncanny valley, especially if paired with the result of cluster 1. Indeed, it seems to suggest that we tend to evaluate positively those robots that are the most similar to us but when their humanlikeness mismatches our mental model of humanlikeness, this positive evaluation turns negative (cluster 1).

4.1.2 Communion and Agency: Perceived Gender and Appearance (RQ1b and RQ1d). The results of PM1 and PM2 disclosed that the **robot's perceived gender did affect the activation of stereotypes but only towards communion, and not for gender neutrality** (**RQ1b**). Indeed, while the perception of agency was mainly affected by the robot's humanlikness, the perception of communion was influenced by both the robot's perceived masculinity (negatively) and its perceived femininity (positively). This result might be taken to support the hypothesis of a default male gender for robots [36, 37]. Indeed, if the hypothesis that humanoid robots are male by default was true, agency, a predominantly male trait, would not need to be regulated by masculinity, as it could directly be regulated by humanlikeness. However, communion, a predominantly female trait, would still need to be regulated by femininity and, by contrast, by masculinity, as it could not directly be regulated by humanlikeness, which is precisely what happens here

In terms of appearance features (**RQ1b** and **RQ1d**), **a full body (high BM), especially if interiorized (high SL), aids the perception of agency**, in particular when coupled with humanlikeness, but **the presence of BM (e.g., arms, hands, torso) intrudes the perception of communion**, both when paired with masculinity and with femininity. These results seem to suggest that BM regulate which robots are perceived as humanlike and hence agentic, and, by extension, which robots are perceived as masculine and hence not communal. Interestingly though, they also seem to suggest that BM regulate which robots are perceived as *not* feminine and hence *not* communal. The identification of what is feminine and communal by opposition to what is masculine and agentic supports once more the hypothesis of a default male gender for robots, and the long-standing tradition of female markedness starting with De Beauvoir [12]: female gender(edness), both in humans and robots, is not a gender *per se*, rather it is a variation of male gender(edness).

SL do not have the same power as BM. Yet, the fact that SL have a role (albeit marginal) in eliciting communion might be taken to support our initial claim that the very features used to imbue gender into robots might activate stereotypes. SL features refer to appearance cues, socio-culturally, stereotypically associated with women (e.g., eyelashes, apparel) [37]. Hence, it does not surprise that they affect the attribution of communion, a trait socio-culturally, stereotypically associated with women, to female robots.

4.1.3 Communion and Agency: Perceived Age and Appearance (RQ1c and RQ1d). Unsurprisingly, judging from our inspection of the clusters in Section 4.1.1, the results of PM3 reveal that **a robot's**

perceived age has a strong impact on the activation of stereotypical traits, both in terms of communion (negative) **and agency** (positive, **RQ1c**), and that **FF (e.g., head, eyes) and SL (e.g., skin, head hair) affect this activation even further** (**RQ1c** and **RQ1d**). The literature on age stereotypes only rarely covers stereotypes affecting children and adolescents [20]. When these stereotypes are investigated though, children and adolescents are usually viewed as high in communion and low in agency. PM3 and the cluster analysis in Section 4.1.1 confirm this view also for robots. Indeed, they reveal that the *older* a robot is perceived to be, the more agency and less communion it is attributed. At a first glance, this result might seem to contradict the stereotype of elderly people, and by extension oldlike robots, as high in communion and low in agency. Yet, the reader should be aware of the fact that the number of oldlike robots in our study was extremely low (only 3 robots with a perceived age > 60), hence "older" mostly means adultlike.

Given the positive relationship between SL and age, their indirect negative effect on communion does not surprise. Yet, the fact that FF have a negative indirect effect on agency via perceived age, and a positive indirect effect on communion via the same, is particularly compelling. Indeed, this result can be taken to support our initial claim that the features used to imbue age into robots might have a role in activating age stereotypes, especially when tapping into the baby schema. Indeed, childlikeness often makes itself evident on the face of a robot with a bulgier head, disproportionately big with respect to the rest of the body, and big "cartoony" eyes [26].

4.2 Stereotypical Tasks

4.2.1 Task Suitability: Stereotype Content, Structure, and Predictors (RQ2a-d). The results of the second cluster analysis produced only two clusters of robots perceived as low vs. high in task suitability for both male and female tasks (**RQ2a**). This suggests that **there is no ambivalence in task stereotypes**, and that **robots are not usually perceived as suited for female tasks while unsuited for male tasks**, and vice versa. The two clusters we identified, however, seem to keep trace of the task genderedness, as the robots in the high suitability cluster were globally perceived as more suited for male tasks, while those in the low suitability cluster as more suited for female tasks. This result seems to suggest that low and high task suitability are already gendered - *the capability to carry out tasks being male* - and hence do not need further gender labels.

If we take a look at each cluster, what we immediately notice is that the low suitability cluster features half-bodied or sitting robots (e.g., Mero3, Chico), and childlike robots (e.g., Kaspar). On the opposite, the high suitability cluster features full-bodied robots (e.g., Asimo, Romeo). Besides telling us something about people's expectations of the capabilities of full-bodied, humanlike robots, these observations also anticipate the results of PM4-PM6. Indeed, these PMs disclose that **BM play a substantial role in determining task suitability, while perceived gender does not** (**RQ2b**). Moreover, **they show a negative effect of FF on task suitability (PM6), fully mediated by the robot's perceived age** (**RQ2c** and **RQ2d**). From a purely functional standpoint, task suitability should be improved rather than impaired by FF. Hence, this result suggests that FF are taken by participants for their aesthetic, rather than functional, value, and supports once more our claim that a robot's stereotypical appearance has a role in eliciting stereotypes.

4.3 Individual Characteristics

The results of the multilevel regressions disclosed that **participants' age, tendency to anthropomorphize, benevolent and hostile sexism are associated with stereotyping** (**RQ1e** and **RQ2e**). Male gender, hostile and benevolent sexism seem to make participants especially sensitive to male stereotypes, with men perceiving robots as less suited for male tasks, hostile sexism reducing robot's perceived agency, and benevolent sexism, in line with its proverbial paternalism, increasing said agency. Tendency to anthropomorphize and age, instead, increase stereotype ratings irrespective of the stereotypes' genderedness. What these results tell us is that the individual characteristics that predict people's proneness to stereotype robots are the same that predict their proneness to stereotype humans. Agency was shown to predict both suitability for female and suitability for male tasks, while communion only suitability for female tasks. We can thus assume that, in the HRI context, *communion* is the perceptual affordance that suggests a robot's sociability, while *agency* the perceptual affordance that suggests a robot's goal-orientedness and capability to perform tasks.

4.4 Limitations

While this study produced relevant insights, it did not come without limitations. The first limitation was the robot's gender distribution, with more male robots than female ones. Unfortunately, this issue is endemic to the HRI field, as the distribution of robots' perceived gender in the ABOT database is itself skewed [37]. Another limitation of the study has to do with the use of static images. Static images often lack details relevant for the social perceptions of robots, such as the robot's size. While images are good for gaining insights regarding broad HRI phenomena, their findings cannot be taken to extend to actual interactions with robots, and hence need to be replicated under different conditions. A last limitation stems from the lack of investigation of cultural differences. Since the word robot is masculine in some languages, while neutral in others, some participants might have been more prone to stereotyping.

5 CONCLUSIONS

This study provides a clear understanding of the content, structure, and predictors of robot's stereotypes, and shows preliminary evidence of a causal relationship between robot's appearance cues and stereotype activation mediated by certain social categories. Future work should focus on testing the relationship between the content of these stereotypes and the affective reactions and discriminatory tendencies robots elicit. To facilitate this process, we make the dataset we collected, the STereotypes in ROBOts (STROBO) dataset publicly available, to the HRI community (see Appendix 3).

ACKNOWLEDGMENTS

This work is part of the research programme Ethics of Socially Disruptive Technologies (ESDiT), which is funded through the Gravitation programme of the Dutch Ministry of Education, Culture, and Science and the Netherlands Organization for Scientific Research (NWO grant number 024.004.031).

REFERENCES

[1] Andrea E Abele, Naomi Ellemers, Susan T Fiske, Alex Koch, and Vincent Yzerbyt. 2021. Navigating the social world: Toward an integrated framework for evaluating self, individuals, and groups. *Psychological Review* 128, 2 (2021), 290.

[2] Andrea E. Abele and Bogdan Wojciszke. 2007. Agency and Communion From the Perspective of Self Versus Others. *Journal of Personality and Social Psychology* 93, 5 (11 2007), 751–763.

[3] Simone Alesich and Michael Rigby. 2017. Gendered Robots: Implications for Our Humanoid Future. *IEEE Technology and Society Magazine* 36, 2 (6 2017), 50–59.

[4] Jasmin Bernotat, Friederike Eyssel, and Janik Sachse. 2017. Shape It – The Influence of Robot Body Shape on Gender Perception in Robots. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 10652 LNAI. 75–84.

[5] Jasmin Bernotat, Friederike Eyssel, and Janik Sachse. 2019. The (Fe)male Robot: How Robot Body Shape Impacts First Impressions and Trust Towards Robots. *International Journal of Social Robotics* 13, 3 (2019), 477–489.

[6] Dave Berque, Hiroko Chiba, Michiko Ohkura, Peeraya Sripian, and Midori Sugaya. 2020. Fostering cross-cultural research by cross-cultural student teams: A case study related to kawaii (Cute) robot design. In *International Conference on Human-Computer Interaction*. Springer, 553–563.

[7] Piercosma Bisconti and Giulia Perugia. 2021. How Do We Gender Robots? Inquiring the Relationship Between Perceptual Cues and Context of Use. In *Proceedings of GenR 2021 Workshop on Gendering Robots: Ongoing (Re) configurations of Gender in Robotics*. 1–4.

[8] De'Aira Bryant, Jason Borenstein, and Ayanna Howard. 2020. Why should we gender? the effect of robot gendering and occupational stereotypes on human trust and perceived competency. *ACM/IEEE International Conference on Human-Robot Interaction* (2020), 13–21.

[9] Natalia Calvo-Barajas, Giulia Perugia, and Ginevra Castellano. 2020. The effects of robot's facial expressions on children's first impressions of trustworthiness. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 165–171.

[10] Namok Choi, Dale R Fuqua, and Jody L Newman. 2009. Exploratory and Confirmatory Studies of the Structure of the Bem Sex Role Inventory Short Form With Two Divergent Samples. *Ed. and Psychol. Measurement* 69, 4 (2009), 696–705.

[11] Amy J.C. Cuddy, Susan T. Fiske, and Peter Glick. 2007. The BIAS Map: Behaviors From Intergroup Affect and Stereotypes. *Journal of Personality and Social Psychology* 92, 4 (2007), 631–648.

[12] Simone De Beauvoir. 2014. The second sex. In *Classic and Contemporary Readings in Sociology*. Routledge, 118–123.

[13] Alice H Eagly and Anne M Koenig. 2021. The vicious cycle linking stereotypes and social roles. *Current Directions in Psychological Science* 30, 4 (2021), 343–350.

[14] Alice H Eagly and Wendy Wood. 2012. Social role theory. *Handbook of theories of social psychology* 2 (2012).

[15] Nicholas Epley, Adam Waytz, and John T. Cacioppo. 2007. On Seeing Human: A Three-Factor Theory of Anthropomorphism. *Psychological Review* 114, 4 (2007), 864–886.

[16] Friederike Eyssel and Frank Hegel. 2012. (S)he's Got the Look: Gender Stereotyping of Robots. *Journal of Applied Social Psychology* 42, 9 (2012), 2213–2230.

[17] Friederike Eyssel, Dieta Kuchenbrandt, Frank Hegel, and Laura De Ruiter. 2012. Activating elicited agent knowledge: How robot and user features shape the perception of social robots. In *2012 IEEE RO-MAN: The 21st IEEE international symposium on robot and human interactive communication*. IEEE, 851–857.

[18] Susan T Fiske. 1998. Stereotyping , prejudice , and discrimination. In *The handbook of social psychology*. 357–411.

[19] Susan T Fiske. 2012. Managing ambivalent prejudices: smart-but-cold and warm-but-dumb stereotypes. *The Annals of the American Academy of Political and Social Science* 639, 1 (2012), 33–48.

[20] Susan T Fiske. 2017. Prejudices in cultural contexts: Shared stereotypes (gender, age) versus variable stereotypes (race, ethnicity, religion). *Perspectives on psychological science* 12, 5 (2017), 791–799.

[21] Susan T. Fiske. 2017. Prejudices in Cultural Contexts: Shared Stereotypes (Gender, Age) Versus Variable Stereotypes (Race, Ethnicity, Religion). *Perspectives on Psychological Science* 12, 5 (2017), 791–799.

[22] Susan T Fiske, Amy JC Cuddy, and Peter Glick. 2007. Universal dimensions of social cognition: Warmth and competence. *Trends in cognitive sciences* 11, 2 (2007), 77–83.

[23] Susan T Fiske, Amy J.C. Cuddy, Peter Glick, and Jun Xu. 2002. A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology* 82, 6 (2002), 878–902.

[24] Peter Glick and Susan T. Fiske. 1996. The Ambivalent Sexism Inventory: Differentiating Hostile and Benevolent Sexism. *Journal of Personality and Social Psychology* 70, 3 (1996), 491–512.

[25] Melanie L Glocker, Daniel D Langleben, Kosha Ruparel, James W Loughead, Ruben C Gur, and Norbert Sachser. 2009. Baby schema in infant faces induces cuteness perception and motivation for caretaking in adults. *Ethology* 115, 3 (2009), 257–263.

[26] Stephen Jay Gould. 2008. A biological homage to Mickey Mouse. *Ecotone* 4, 1 (2008), 333–340.

[27] Stefano Guidi, Latisha Boor, Laura van der Bij, Okke Rikmenspoel, Robin Foppen, and Giulia Perugia. 2022. Ambivalent Stereotypes Towards Gendered Robots: The (Im)mutability of Bias Towards Female and Neutral Robots. In *International Conference on Social Robotics*. Springer, In Press.

[28] Quirien RM Hover, Ella Velner, Thomas Beelen, Mieke Boon, and Khiet P Truong. 2021. Uncanny, sexy, and threatening robots: The online community's attitude to and perceptions of robots varying in humanlikeness and gender. In *Proceedings of the 2021 ACM/IEEE Int'l Conference on Human-Robot Interaction*. 119–128.

[29] Konrad Lorenz. 1943. The innate forms of potential experience. *Z Tierpsychol* 5 (1943), 235–409.

[30] Konrad Lorenz and Robert Martin. 1971. Studies in animal and human behaviour. *British Journal for the Philosophy of Science* 22, 1 (1971).

[31] Aaron Marcus, Masaaki Kurosu, Xiaojuan Ma, and Ayako Hashizume. 2017. *Cuteness engineering: Designing adorable products and services*. Springer.

[32] Hannah Mieczkowski, Sunny Xun Liu, Jeffrey Hancock, and Byron Reeves. 2019. Helping not hurting: Applying the stereotype content model and BIAS map to social robotics. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 222–229.

[33] Sven Y Neuteboom and Maartje de Graaf. 2021. People's Perceptions of Gendered Robots Performing Gender Stereotypical Tasks. In *International Conference on Social Robotics*. Springer, 24–35.

[34] Hiroshi Nittono. 2022. The Psychology of" Kawaii" and Its Implications for Human-Robot Interaction. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction*. 3–3.

[35] Hiroshi Nittono, Michiko Fukushima, Akihiro Yano, and Hiroki Moriya. 2012. The power of kawaii: Viewing cute images promotes a careful behavior and narrows attentional focus. *PloS one* 7, 9 (2012), e46362.

[36] Oronzo Parlangeli, Paola Palmitesta, Margherita Bracci, Enrica Marchigiani, and Stefano Guidi. 2022. Gender role stereotypes at work in humanoid robots. *Behaviour & Information Technology* 0, 0 (2022), 1–12.

[37] Giulia Perugia, Stefano Guidi, Margherita Bicchi, and Oronzo Parlangeli. 2022. The Shape of Our Bias: Perceived Age and Gender in the Humanoid Robots of the ABOT Database. In *Proceedings of the 2022 17th ACM/IEEE International Conference on Human-Robot Interaction. ACM/IEEE.*

[38] Giulia Perugia and Dominika Lisy. 2022. Robot's Gendering Trouble: A Scoping Review of Gendering Humanoid Robots and its Effects on HRI. *arXiv preprint arXiv:2207.01130* (2022).

[39] Giulia Perugia, Alessandra Rossi, and Silvia Rossi. 2021. Gender Revealed: Evaluating the Genderedness of Furhat's Predefined Faces. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 13086 LNAI. Springer International Publishing, 36–47. 4

[40] Elizabeth Phillips, Xuan Zhao, Daniel Ullman, and Bertram F Malle. 2018. What is human-like?: Decomposing robots' human-like appearance using the anthropomorphic robot (abot) database. In *2018 13th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 105–113.

[41] Diane Ponterotto. 2016. Resisting the male gaze: feminist responses to the" normatization" of the female body in Western culture. *Journal of international women's studies* 17, 1 (2016), 133–151.

[42] Daniel J Rea, Yan Wang, and James E Young. 2015. Check your stereotypes at the door: an analysis of gender typecasts in social human-robot interaction. In *International conference on social robotics*. Springer, 554–563.

[43] Byron Reeves, Jeff Hancock, and Xun Liu. 2020. Social robots are like real people: First impressions, attributes, and stereotyping of social robots. *Technology, Mind, and Behavior* 1, 1 (2020), 1–37.

[44] Natalia Reich-Stiebert and Friederike Eyssel. 2017. (Ir)relevance of Gender? on the Influence of Gender Stereotypes on Learning with a Robot. In *2017 12th ACM/IEEE Int'l Conference on Human-Robot Interaction (HRI*. IEEE, 166–176.

[45] Jennifer Robertson. 2010. Gendering humanoid robots: Robo-sexism in Japan. *Body and Society* 16, 2 (2010), 1–36.

[46] Chiara Rollero, Peter Glick, and Stefano Tartaglia. 2014. Psychometric properties of short versions of the Ambivalent Sexism Inventory and Ambivalence Toward Men Inventory. *TPM - Testing, Psychometrics, Methodology in Applied Psychology* 21, 2 (2014), 149–159.

[47] Katta Spiel, Oliver L Haimson, and Danielle Lottridge. 2019. How to do better with gender on surveys: a guide for HCI researchers. *Interactions* 26, 4 (2019), 62–65.

[48] Joachim Stoeber. 1999. The Social Desirability Scale-17 (SDS-17): Development and first results on reliability and validity. *Diagnostica* 45, 4 (1999), 173–177.

[49] Adam Waytz, John Cacioppo, and Nicholas Epley. 2010. Who sees human? The stability and importance of individual differences in anthropomorphism. *Perspectives on Psychological Science* 5, 3 (5 2010), 219–232.