



Investigating the Impact of Gender Stereotypes in Authority on Avatar Robots

Yuan-Chia Chang
chang.yuanchia.62s@st.kyoto-u.ac.jp
Kyoto University
Japan

Daniel J. Rea
daniel.rea@unb.ca
University of New Brunswick
Canada

Takayuki Kanda
kanda@i.kyoto-u.ac.jp
Kyoto University
Japan

ABSTRACT

We investigate how gender stereotypes in authority influence the perceptions and behavior of avatar robots operators and their interlocutors. Gender stereotypes, which typically place men in more authoritative positions than women, are present in not only inter-human but also human-robot interaction. As avatar robots become more integrated into our lives and serve for diverse usages, they may be utilized in positions where they require authority. We study how avatar robot gender and operator gender affect expressions and perception of gender stereotypes in a customer service scenario with 41 pairs of participants. Operators controlled binary gendered avatar robots one at a time, acting as shopkeepers that had to assert authority over customers behaving improperly. The operators perceived their authority to be higher with male avatar robots compared to female ones, regardless of operator gender. We did not detect an effect on customer's perception of the shopkeeper's authority. While less than half of operators and customers perceived authority for reasons related to traditional gender stereotypes, others observed behaviors that did not align with stereotypes. Avatar embodiment may also help operators assert authority safely due to being physically hidden from the customers.

CCS CONCEPTS

• Human-centered computing → Empirical studies in HCI.

KEYWORDS

Human-robot interaction; gender stereotype; authority; avatar; avatar robot

ACM Reference Format:

Yuan-Chia Chang, Daniel J. Rea, and Takayuki Kanda. 2024. Investigating the Impact of Gender Stereotypes in Authority on Avatar Robots. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction (HRI '24)*, March 11–14, 2024, Boulder, CO, USA. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3610977.3634985>

1 INTRODUCTION

The influence of gender stereotypes on authority is rooted in many cultures [86] for decades. Traditionally, men are put in positions

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](https://permissions.acm.org).
HRI '24, March 11–14, 2024, Boulder, CO, USA

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-0322-5/24/03...\$15.00
<https://doi.org/10.1145/3610977.3634985>

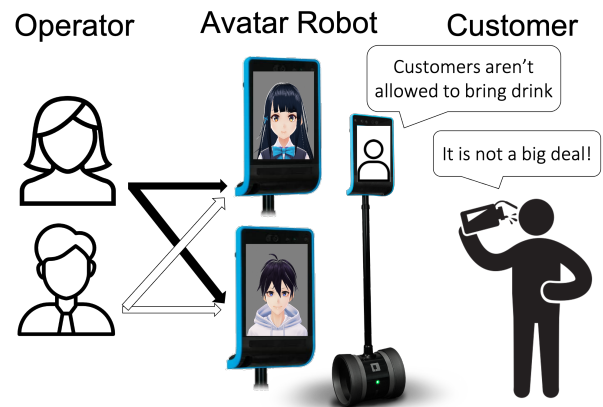


Figure 1: People embodied different gendered avatar robot shopkeepers to assert their authority in our customer service scenario. Icon by Flaticon.com (<https://www.flaticon.com/>)

of power and influence, thus occupy more authoritative roles than women [3, 15, 54, 87]. Women usually express less authority in their job [54, 86], and occupy fewer authoritative roles. While gender equality has gained global recognition, leading to increased perceptions of female competence in some Western countries [27], gender stereotypes, especially concerning women, have evolved more rapidly than those about men [51]. However, in some Eastern countries like Japan, women still face traditional stereotyped expectations [11]. Androgynous and non-binary people also face negative stereotypes [40]. These stereotypes impact both how people perceive and are perceived by others.

Stereotypes are also present in human-agent or human-robot interaction [19, 29, 58, 64, 81, 82]. Robots perceived as male are seen as more dominant, having more agency, and being more suitable to be security guard robot, an authority role. Robots perceived as female are seen as more communal, and more suitable to healthcare than male robots [29, 64, 81]. Even robots with gender neutral designs can activate communal stereotypes in men and agency stereotypes in women [35]. However, stereotypes may be less pronounced in robots than in humans [67]. Still, this evidence demonstrates that stereotypes affect perceptions and interactions with robots, even when aiming for a gender neutral design.

Avatars and avatar robots, representations of human operators, have already been deployed in public, in both virtual and the physical worlds. In the physical world, they used to help increase accessibility [79], and improving business places [1, 4]. As the appearance of an avatar (e.g., the graphical or robot design) is changeable, they enable operators to embody characteristics that are different from their real selves, such as age [10], gender [32, 50], race [8, 34, 63], species [6] and mobility [17, 79]. Avatar embodiment can alter the

perceptions of and toward operators, such as helping reduce gender dysphoria or experience more diversity [55, 70].

People already change characteristics like appearance or gender in virtual spaces like online games or the VTuber industry. These spaces regularly see interactions with real world stereotypes, such as men acting in stereotyped ways when embodying female avatars [20], made more common by people preferring to play female avatars [21]. While women are less likely to engage in gender swapping [41], they sometimes do it to avoid stereotypes in male dominated areas [42]. In the physical world, people also embody different genders, such as portraying mascots of different genders or by use of an avatar robot. The mere ability to adjust gender can prime people to think more of gender-relevant traits and engage with stereotypes. Whether this makes people accept gender stereotypes [20] or rebel against them [45, 77] is a complex issue influenced by a person's own thoughts and experiences with various stereotypes in the world. While people's performance and mental workload in cognitive and social tasks [73] would be affected under stereotype threat [60]. These reactions to stereotypes can even be activated unconsciously via subtle cues [9, 23, 45].

Both now and in the future, avatar robots could influence how people use, perceive, and experience stereotyped behavior in the physical world. We study how people who embody avatars of the both same and different genders are perceived and express themselves within the context of traditional binary gender stereotypes in authority. We observe if an avatar's gender enhances or inhibits stereotypes, considering the potential of avatar robots in mitigating gender stereotypes.

2 RELATED WORK

2.1 Gender Stereotype in Authority

Gender stereotypes have been criticized for decades, and the gender gap in authority is one indicator of the phenomenon. Traditionally, roles of men and women have been assigned to occupational and domestic roles respectively [11, 28, 72]. Men are more likely to occupy high-level roles than women [3, 15, 26, 54, 87], which leads to their higher perceived authority and social influence [22, 25, 26]. Consequently, women typically express less authority in their working space [86], and have smaller occupational networks than men [15]. Such stereotypes also appear in human-robot interaction. A robot's persona can be easily manipulated by many cues (such as appearance [29], voice [65], demeanor and names [68]) which will influence the human's behavior toward them and what information we can elicit from them [66]. Even subtle cues can affect people's expectations and stereotypes of a robot [29].

On the other hand, work related to robot authority is still limited [5]. Though some works use robots with low-pitched voices [5] and male appearances [7] to assert authority, gender stereotypes are not their focus. How gender stereotypes affect the perception of a robot's authority is a topic that requires further exploration.

2.2 Stereotype and Avatar Self-Representation

When people embody avatars, their behavior is influenced by their mental models of both themselves and their virtual avatars [88]. For example, people act more intimately and more confidently when using more attractive avatars [53, 88]. These effects can be found

with different embodiment types such as static pictures [46, 83], video-based virtual environment (e.g., video games) [12, 53] or virtual reality [8, 32–34, 50, 57, 63]. From this, research has explored making people embody characteristics that differ from their own to see how it affects their stereotypes. They found stereotypes applied to avatars are influenced by the interplay between embodiment and context. For example, with a darker-skinned embodiment, the avatar experience increased racial bias in job interview scenarios [34], but decreased racial bias in exercise scenario [8]. Similar results have also been found for gender stereotypes [12, 32, 50].

Avatars can help mitigate stereotypes by helping people experience stereotypes not typically applied to their real appearance [12, 32]. Experiencing negative experiences with gender stereotypes can reduce implicit gender bias [12], such as reducing men's biases against woman in STEM [32]. The change in stereotypes varies on the basis of the scenario and embodiment, and there is yet no specific strategy to reduce stereotypes generally.

2.3 Telepresence Robot and Avatar Robot

Telepresence robots and avatar robots facilitate remote interactions and enable people to control the robots and engage in physical activities. Telepresence robots, which are equipped with a screen displaying the operator's face and mobility capabilities, enhance the presence of remote users [13]. They are particularly suited for situations where the operator's identity is important [48], such as participating in academic conferences [59] or in elderly care settings [31]. On the other hand, as robots capable of remote operation can function as avatar robots without the need for a screen to display the operator's face, they can effectively conceal the operator's actual physical or mental state, such as race, gender or conditions like bedridden status [79] or social anxiety disorder [89].

Currently, the studies about avatar and avatar robots focus on the impact of the people who embody different avatars. However, there is a lack of understanding of how other people perceive avatar robots specifically and how both the operators and interlocutors influence each other, in terms of their authority. Given the deployment of avatar robots in daily life, understanding real interpersonal interactions with gendered avatar robots may help us gain insights to mitigate gender stereotypes.

3 EXPERIMENT

Our study focused on observing expressions and perceptions of gender stereotypes related to authority when participants embody different gendered avatar robots. To study both the expressions and reactions, we recruited participants for both an operator and an interlocutor for the interaction. They participated in a role-play scenario set in a shop where the interlocutor (customer) challenged the authority of the avatar robot operator (shopkeeper) by doing prohibited behavior against the shop's rules. This is distinct from situations with legitimate figures of authority like police officers or security guards, who are often obeyed with less resistance.

3.1 Hypothesis

Previous research discussing traditional gender stereotype in authority showed men are usually perceived as more authoritative than women. Thus, the following hypothesis were proposed:

H1. Authority:

H1-1. Customers will perceive a male avatar robot shopkeeper as having higher authority than a female avatar robot shopkeeper.

H1-2. Operators (shopkeepers) will perceive higher self-authority when operating a male avatar robot shopkeeper than a female avatar robot shopkeeper.

As customers should tend to perceive a male avatar as having more authority, they may be more inclined to give up making unreasonable requests to it sooner. Thus, H2 is proposed.

H2. Time: Customers will spend less time arguing with a male avatar robot shopkeeper than a female avatar robot shopkeeper.

As people’s workload would be affected by stereotypes [60], we measured operator’s overall workload.

H3. Work load: Operators (shopkeepers) will have less work load when they operate a male avatar robot shopkeeper than a female avatar robot shopkeeper.

3.2 Conditions

We conducted a 2 x 2 study (avatar gender x operator gender) in mixed method design.

- Within-subject factor: avatar gender (male and female)
- Between-subject factor: operator gender (male and female)

The order of avatar gender was counterbalanced.

3.3 Scenario

3.3.1 Role-play method. As participants can engage with specific scenarios for the experiment and leverage their daily experiences, role-play method is widely utilized [24, 49, 56]. Since using robots to interact with real, difficult customers presents ethical dilemmas for shopkeepers and shops, we adopted this method in our study.

3.3.2 The role of a customer. The customer selfishly argued with the shopkeeper in order to keep drinking in the shop. The task continued until the customer gave up arguing or convinced the shopkeeper. There was no minimum time to stay in the store, but the experimenter ended the study after 30 minutes. This rule was not revealed in advance to prevent participants thinking about a “correct” amount of time to stay in the room. For training, customers watched a video depicting a real-life situation where a difficult customer argued with a station attendant.

3.3.3 The role of an operator. The operators used two different gendered avatars, one per condition. They were told to behave according to the avatar’s gender (to make interaction more natural as Japanese language features gendered grammar), but we did not provide guidance on what appropriate gender behavior is. While this could still unconsciously activate gender stereotypes [9, 23, 45], as discussed in the introduction, our scenario is similar to current real world avatar applications, and it is unclear if this activation will enhance or mitigate stereotypes.

Shopkeepers needed to assert their shopkeeper authority to prevent customers from bringing drinks indoors. They could not resort to other people like security guards or managers, as we did not want other authority figures involved in the interactions.

3.4 Avatar Robot



Figure 2: Four out of the 22 female avatars and 4 out of 6 male avatars used in the experiments. For each gender, the left three were the most frequently selected by operators. The fourth helps show the avatar diversity in our pool. Avatars were from booth.pm (<https://booth.pm/ja>)

3.4.1 Avatar appearance. Our avatar robot was a Double 3 robot that displayed the face of a 3D avatar on its screen (see the middle of Fig. 1). To avoid our bias in avatar design, we applied objective criteria to select avatars from an avatar website (booth.pm) for a diverse pool: (1) original full-body 3D human characters (about 10% filtered) (2) no sci-fi elements like animal features, suitable for shopkeepers (46% filtered) (3) modifiable in VRoid Studio (.vroid file) (40% filtered) (4) permissible for beyond personal use (2% filtered). We note that our criteria may not result in all avatars that meet the requirements as the avatar models are not obliged to have accurate keywords. This process yielded 22 female and 6 male avatars from approximately 1000 (as of January 2023, see Fig. 2), reflecting a public preference for female avatars [21], despite the gender imbalance.

3.4.2 Avatar voice. Voice changer software was used to ensure the operators’ voices better matched the avatar gender in both conditions for control. There are 8 male and 9 female voice options, each with varying timbres (e.g., vigorous men voice, voice like young lady or middle-age men), and they can be roughly categorized into low, middle, and high pitches. It is noted that the voice change depends on the participant’s own voice, so different people using the same voice changer setting may sound different.

3.4.3 Synchronizing operator with avatars. Operators’ facial expressions were captured and mapped to the avatars displayed on the Double 3’s screen. We used the Double 3’s web app to communicate through and move the robot. Operators could see their avatars in the upper-left corner of the interface and hear the changed voice in the headphones for increased awareness of the avatar gender.

3.5 Method

3.5.1 Environment. The experiment took place in a room decorated like a hat shop, with several hats and two mirrors in the room (Fig. 3). A sign saying that eating and drinking was prohibited was put in the entrance to make sure it could be seen easily.

3.5.2 Participants. During recruitment, people had the option to identify their gender as male, female, and prefer not to say. Very few people did not disclose their gender identity, reflecting the known cultural difficulty in Japan of openly discussing non-binary or changed identities [80]. Therefore, eighty-eight Japanese participants, who identified as male or female were recruited, and with

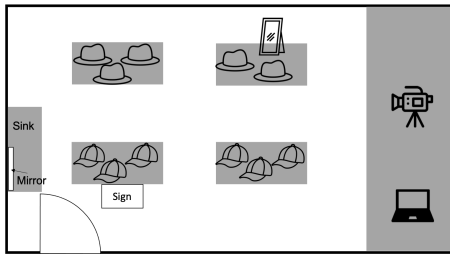


Figure 3: The settings of the experiment room. Icon by Flaticon.com (<https://www.flaticon.com/>)

varied ages as stereotypes can vary with age [27]. They were assigned to the roles of operators and customers randomly to form 44 pairs, with customer gender balanced across pairs as a control. We excluded 3 pairs. Two pairs did not follow instructions (e.g., not arguing about rules), and one participant in a group wanted to withdraw. Thus, we finally had 41 groups, with 20 female operators (ages 21 to 64, $M=39.10$, $SD=13.43$) and 21 male operators (ages 18 to 56 $M=30.71$, $SD=13.56$); 21 female customers (ages 19 to 64 $M=37.68$, $SD=13.9$), and 20 male customers (ages 18 to 64 $M=38.55$, $SD=15.11$). The experiment took 2 hours, and participants were compensated with 3000 JPY.

3.5.3 Procedure. The operators and customers were directed to separate rooms so they were not aware of each other. After signing the consent form, they were instructed about their roles (see section 3.3). Then, operators learned how to operate the robot and select the avatar. They viewed all avatars in a scrollable window and explored voice options freely, but were limited to using avatars and voices that matched the condition (see section 3.2). The session started when the customers entered the room with the avatar robot, which was controlled by operators remotely. It concluded when customers left the room. Then operators and customers filled out their questionnaires. After both conditions, we conducted a semi-structured interview with each participant. Our study received Institutional Review Board approval.

3.5.4 Measurements. The questionnaire items (in Japanese) were scored with 7-point Likert items. The operators were asked:

- **Self-authority:** Despite the existence of various types of authority with validated measurement scales [69, 71], they are unsuitable for measuring a robot shopkeeper’s authority. We modified the single-item questionnaire used in [39], which assesses the authority perception of a robot placed in the same room with participants, to measure the authority associated with the presence of our shopkeeper robot and which was phrased as “How much authority did you have when operating the robot?”.
- **Work load:** As the NASA-TLX is a standard measure in teleoperation [78], the Japanese version [36, 37] was used, and it was modified to a 7-point scale.

The customers were asked:

- **Shopkeeper’s perceived authority:** a single question modified from [39] which was phrased as “How much authority did you feel the shopkeeper had?”.
- **Manipulation check:** a single question directly asked “What gender characteristics did you think of the shopkeeper robot as having?”, the same as the strategy in [29]. Lower numbers indicated feminine and higher numbers indicated masculine.

We recorded the experiment time for verifying H2.

- **Time:** the time the customer spent in the shop.

4 RESULTS

4.1 Manipulation Check

We asked customers to rate the gender characteristics of avatars to verify the alignment with our experimental conditions. We found the results were non-normal (Shapiro-Wilk, $W=.823$, $p<.001$), and had non-equal variances (Levene’s test, $p<.05$). A Wilcoxon Rank Sum Test found a significant effect of avatar gender ($Z=7.22$, $p<.001$), indicating that customers felt the avatars in female avatar condition was more feminine ($M=1.68$, $SD=1.29$) than avatars in male avatar condition ($M=5.56$, $SD=1.45$).

4.2 Observations

Shopkeepers (operators) usually began by politely asking customers to follow the shop rules prohibiting eating and drinking in the shop. As customers needed to play the role of a selfish customer, they started arguing with the shopkeepers. Some of them even mimicked exaggerated drinking to make them perceived as more selfish. Meanwhile, customers browsed the hats while shopkeepers continued asking the customers to leave. In general, we saw varied and complex behavior from customers and operators, though they both exhibited behavior patterns. Avatar and customer genders are labeled as ff (female-female), fm (female-male), mf (male-female), and mm (male-male). The number following each label indicates the count of tasks exhibiting that pattern.

4.2.1 Customer patterns:

Customers adopted the following behaviors to the shopkeepers.

Making excuses (ff:20, fm:20, mf: 20, mm: 20):

- **Customer:** I will not spill the drink or make a mess. It’s really hot outside. I’d prefer to stay indoors.
- **Shopkeeper:** There’s still a chance of spilling, and we have rules that prohibit drinking inside. Could you please go outside?
- **Shopkeeper:** May I take your drink while you are in the store?
- **Customer:** I am just holding it, not drinking. Is it not allowed?

Ignoring (ff:10, fm:6, mf: 8, mm: 6): Some customers ignored the shopkeeper’s requests. They gave casual feedback and continued their inappropriate behavior. For example, when the shopkeeper said “Excuse me, drinking is prohibited here.”, the customer replied:

- **Customer:** Eh? Really? (Continues browsing the hats)
- **Customer:** I will leave soon. (Continues drinking, see Fig. 4)

Showing anger (ff:5, fm:3, mf: 5, mm: 1): Some customers were annoyed by the shopkeeper and used strong wording in response.

- **Shopkeeper:** I am sorry, but if you want to have a drink, I believe it would be better to do so outside the store.
- **Customer:** No. Could you be more polite? I am your customer!
- **Shopkeeper:** Drinking is prohibited. Could you please go outside?
- **Customer:** Shut up! (With emotional gestures, see Fig. 5, left)

4.2.2 Shopkeeper patterns:

Typically, shopkeepers kept politely repeating the requests and pointing out the rule of the store.



Figure 4: Customer ignored the shopkeepers and continued selfishly drinking actions.



Figure 5: The customers and shopkeepers engaged in a strong disputation. Left: the customer said “Shut up!” to the shopkeeper. Right: the shopkeeper said “Please get out!” repeatedly and came very close to the customer.

Following and watching the customers (ff:20, fm:17, mf: 20, mm: 17): Shopkeepers kept the customers in their view by moving or turning (see Fig. 6).

Using strong action to command and assert authority (ff:8, fm: 4, mf: 2, mm: 4):

- *Shopkeeper:* Could you please go out?
Customer: No.
Shopkeeper: Please go out! Please go out! Please go out! (Repeated while moving very close to the customer. see Fig. 5, right)
- *Shopkeeper:* Drinking is prohibited, could you please go outside?
Customer: But it is very cold outside.
Shopkeeper: We have the right to choose the customer we want to serve. Please follow the rules.

Most pairs had several loops of the above behaviors during their interactions. Customers left the room when they felt that continuing the conversation would not change the situation.

4.3 Choice of Statistical Test

We analyze our data with a 2x2 (avatar gender x operator gender) mixed ANCOVA, with customer gender as a control variable and operator and customer age as covariates. As the control variable and covariates make visualization difficult, our figures do not include them unless otherwise appropriate.

We found that some of our data was non-normal by Shapiro-Wilk tests, and all of the data passed Levene’s test of homogeneous variances (each test reported later). Based on standard criteria [14, 18], ANOVA is robust to non-normality, meaning that both the normal and non-normal ANOVA have a .025 to .075 Type I error rate for samples from a given population and an alpha of .05 (the Bradley Criterion [18]). This was tested across 1308 conditions, varying equal and unequal group sample sizes, sample sizes, shapes of the distributions, and more, simulated 10000 times with Monte Carlo Simulation - the non-normal ANOVA *always* passed the Bradley



Figure 6: The shopkeeper kept following the customer.

Criteria for robustness to Type I error. [14]. This has been confirmed for Likert-like data [61], with analytic arguments [62], empirical investigations [16], and with additional Monte Carlo simulation approaches [38], including smaller ($N=25$) sample sizes [74]. Because of this robustness and that there is no true non-parametric alternative to a 2 way ANOVA [52], we continued with the test and discuss this in section 6.

4.4 Hypothesis Verification

Customers’ perception of shopkeeper’s authority The data had homogenous variances (Levene’s, $p > .05$) and was non-normal (Shapiro-Wilk, $W = .909, p < .001$). Male operators averagely rated their self-authority as 3.14 (SD= 1.62) when operating male avatar robot and 2.81 (SD= 1.63) for female. Female operators averagely rated their self-authority as 3.45 (SD= 1.79) when operating male avatar robot and 3.05 (SD= 1.76) for female (Fig. 7a).

We did not find main effects of avatar gender ($F_{1,35} = 1.143, p = .292, \eta^2 = .030$) and operator gender ($F_{1,35} < .001, p = .986, \eta^2 = < .001$). There was no interaction effect between avatar and operator gender ($F_{1,35} = .064, p = .802, \eta^2 = .002$). This did not support H1-1.

Operators’ self-authority The data had homogenous variances (Levene’s, $p > .05$), and was non-normal (Shapiro-Wilk, $W = .886, p < .001$). Male operators averagely rated their self-authority as 2.71 (SD= 1.30) when operating male avatar robot and 2.52 (SD= 1.36) for female. Female operators averagely rated their self-authority as 3.25 (SD= 1.77) when operating male avatar robot and 3.05 (SD= 1.43) for female (Fig. 7b).

We found an effect of avatar gender ($F_{1,35} = 5.945, p = .020, \eta^2 = .145$). We did not find an effect of operator gender ($F_{1,35} = .201, p = .657, \eta^2 = .006$). There was no interaction effect between the avatar gender and operator gender ($F_{1,35} = .472, p = .497, \eta^2 = .013$). The significant effect of avatar gender supported H1-2, indicating operators perceived themselves as having higher authority when operating a male avatar robot, regardless of operator gender.

Time The data had homogenous variances (Levene’s, $p > .05$), and was non-normal (Shapiro-Wilk, $W = .845, p < .001$). As the distribution was right-skewed, we transformed it by taking the logarithm, which did not show evidence of non-normality ($W = .97, p = .183$) and had homogenous variances (Levene’s, $p > .05$). Male customers spent averagely 518.24 seconds (SD= 399.24) when interacting with

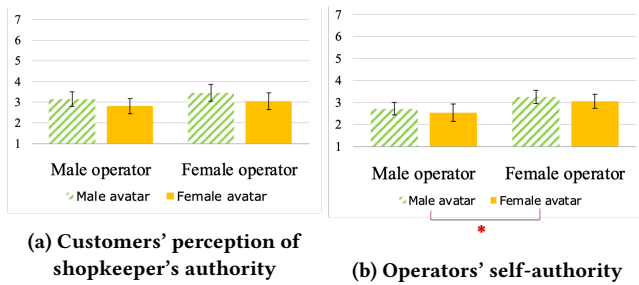


Figure 7: (a) There is no effect in customer’s perception of shopkeeper’s authority. (b) Operators perceived themselves as having significantly higher authority when using male avatars. Error bars show standard error.

the male shopkeeper, and 515.71 seconds (SD= 304.07) for female. Female customers spent averagely 392.00 seconds (SD= 288.77) when interacting with the male shopkeeper, and 449.45 seconds (SD= 368.32) for female (Fig. 8a).

There is no main effect of avatar gender ($F_{1,35} = .026, p = .872, \eta^2 < .001$) and operator gender ($F_{1,35} = .999, p = .324, \eta^2 = .028$). There was no interaction effect between avatar gender and operator gender ($F_{1,35} = .17, p = .682, \eta^2 = .005$). The results did not support H2.

Operators’ work load The data had homogenous variances (Levene’s, $p > .05$), and was normally distributed (Shapiro-Wilk, $W = .985, p = .458$). Male operators rated their work load as 4.16 (SD= 1.09) when operating male avatar robot, and 4.26 (SD= 1.06) for female. Female operators rated their work load as 4.44 (SD= 1.19) when operating male avatar robot, and 4.43 (SD= 1.35) for female. We did not find main effects of avatar gender ($F_{1,35} = .129, p = .722, \eta^2 = .004$) and operator gender ($F_{1,35} = 1.758, p = .193, \eta^2 = .048$). The results did not support H3.

We found an interaction effect between operator gender and customer gender on work load ($F_{1,35} = 4.224, p = .047, \eta^2 = .108$) (see Fig. 8b). A Post-hoc Turkey’s HSD test for multiple comparisons was non-significant for all pairs. However, there was a statistical trend in the difference between the female operator-female customer condition to feel more workload (M= 5.145, SD= 1.186) than the male operator-female customer condition (M= 4.064, SD= .865) ($p = .069, 95\% \text{ C.I.} = [-.112, 2.473]$). The difference between female operator-female customer condition and female operator-male customer condition (M= 3.965, SD= 1.084) was also a statistical trend ($p = .099, 95\% \text{ C.I.} = [-.201, 2.51]$).

4.5 Qualitative Results

We conducted thematic open coding to identify the themes and patterns in post-study interviews. The participant is labelled for operator (o), customer (c), male (m), and female (f). The number following m/f indicates the participants’ age. The themes of how operators perceived their own authority and how customers perceived the shopkeeper’s authority, summarized in the right column of Table 1 and Table 2. The left column shows the groups of the themes for a clearer relation to stereotypes.

4.5.1 How operators perceived their self-authority.

Theme: operators’ own stereotypes. 11 operators believed the male avatar should be more forceful than the female avatar. They

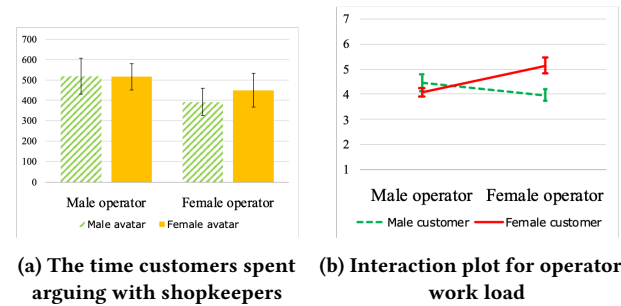


Figure 8: (a) There is no effect in time customers spent arguing with shopkeepers. (b) There was an interaction effect between operator gender and customer gender in operators’ work load. Error bars show standard error.

leveraged male attributes to assert authority, as the male’s deep and intimidating voice was perceived as more authoritative and capable of applying pressure to others. Some operators judged the authority from their experiences with gender stereotypes, not the experiment. They made comments like:

Male are more authoritative than female, but not during the experiment. In society, men decide everything...shop staffs would all be managed by a male supervisor. - o31 (f63)

I did not feel I had authority in the experiment. However, I think males have higher authority from the perspective of the customer due to men’s social image. - o12 (m46)

Theme: Observing the customer’s stereotypical behavior. From observations in the experiment, 7 operators comment like:

I think the male avatar was a little bit more authoritative than the female avatar. The customer tried to argue with the female avatar, but not with the male one. - o17 (f21)

The customer spoke strongly to the female avatar. He wanted the manager to correct her behavior. -o23 (f48)

Theme: Belief in gender quality. Two operators expressed their belief in gender equality, in some sense explicitly acknowledging and working to counter stereotypes. They comment like:

I think everyone is equal. - o15 (f26)

In customer service, the terms should be the same for men and women, so I did not use different terms for male and female avatars. - o41 (m31)

Theme: Observing customer’s behavior. 16 operators did not attribute customer behavior to the influence of stereotypes after observing their actions, but their negotiation skills or familiarity with the tasks.

Theme: Lacking of authority. 5 operators observed that customers’ treatment of the robot said it lacked authority.

4.5.2 How customers perceived shopkeeper’s authority.

Theme: Customer’s own stereotype. 7 customers’ own stereotypes affected their preconception that they should follow male shopkeepers more. Their comments include:

Following the male shopkeeper was easier. In real life, men are more powerful, pressuring and intimidating than women, even when both say the same thing. - c39 (m46)

Table 1: Reasons for operators’ self-authority. Bracketed numbers indicate number of people with each perspective.

stereotype-related (18)	Own stereotypes (11)
	Observing customer’s stereotypical behavior (7)
Unrelated to stereotypes (23)	Belief in gender equality (2)
	Observing customer’s behavior (16)
	Lack of authority (5)

The male avatar’s authority was stronger. From my working experiences in hospitals, patients listen more to males, older, larger body size, and in higher positions. - c30 (f38)

Theme: Observing shopkeeper’s stereotypical behavior. From observations in the experiment, 10 customers comment like:

The male avatar was stronger than the female one, because the male spoke in a commanding way; the female was just asking. - c27 (f39)

The female’s authority seemed low due to her passive way of communicating. The male avatar’s assertive and convincing style made his guidance more effective. - c10 (f24)

Theme: Observing shopkeeper’s behavior. 24 customers did not attribute the reasons to gender stereotype, but all based on their observations of shopkeepers. They either felt avatars assert strong authority through wording, voice, tone, and the direct way of speaking or made similar actions between conditions. Their comments were similar to:

The female avatar had a stronger sense of authority, because she was more pressuring. She was so assertive that I couldn’t really argue. - c8 (f46)

The robots said the same things in the same weak way. There was no difference between conditions. - c36 (f34)

4.5.3 How the embodiment of the avatar robot affected operators.

When asked if the avatar helped them accomplish their tasks, many participants mentioned the avatar robot’s embodiment (both physical and virtual) helped them. 24 operators expressed that the anonymity offered by the avatar robot made them feel secure when giving commands to customers, eliminating concerns about potential confrontation, and even enabling them to use stronger words. Their comments included the following:

Using an avatar help give firm instructions. Hiding my real face and voice, and not being there in person, keeps me safe from potential customer reactions. - o10 (f42)

The customer angrily asked for my name. However, since my face was hidden, if we meet each other on the street, he wouldn’t recognize me, making me feel safe. - o7 (f27)

On the other hand, 14 operators felt using an avatar did not help them accomplish the task. They preferred face-to-face communication, because they can use non-verbal cues such as gestures, facial expression and the nuance to communicate. For example:

I experienced a greater sense of unease compared with face-to-face interactions. The limitations (unable to use non-verbal cues) made it more challenging for me to deal with the situation. - o28 (f46)

Table 2: Reasons for customers’ perception of shopkeeper’s authority. Bracketed numbers indicate how many people shared each perspective.

Stereotype-related (17)	Own stereotype (7)
	Observing shopkeeper’s stereotypical behavior (10)
Unrelated to stereotype (24)	Observing shopkeeper’s behavior (24)

5 DISCUSSION

5.1 Interpretation of the Results

Quantitative results indicated that operators felt more authoritative using male avatars, consistent with traditional gender stereotypes. However, we did not detect an effect for how customers perceived the authority of the shopkeepers. Qualitatively, less than half of operators and customers attributed perceived authority to stereotypes, either interpreting behavior through stereotypes or expressing stereotypical views on them. Operators’ belief in stereotypes may have influenced their self-perceived authority with male avatars, yet this wasn’t evident in their actual customer interactions. Many operators said they were too busy deal with customers to “behave” differently between conditions. This leads to more customers not noticing stereotypical behavior, which might explain the lack of detected impact on customers’ authority perception.

5.2 Effects of Avatar Appearance and Voices

Variations in avatar appearances and voices could lead to different perceptions of authority, gender, and stereotypes. For internal validity, it is important for the perception our avatars to align with the intended gender. We calculated the variance of perceived gender (7-point scale, higher scores indicating more masculinity) for the most selected avatars, focusing on the top three due to their greater participant influence (see Fig. 2). The top male avatars scored 5.27 to 5.83 (voice pitches mix), and females 1.25 to 2.5 (mostly high and middle pitch voices). The scores matched the avatars’ intended genders, but further study is needed to understand how variance in presented gender affects perceived authority.

Given the possibilities in avatar design, we have to be careful about how they influence perceived authority and whether they propagate stereotypes. For instance, avatars with gloomy faces or downcast eyes (perceived as weak [44]) might suggest low authority. Low-pitched voices are seen as more authoritative [5]. Using publicly created avatars revealed a trend towards young, slender bodies and stereotypical features like cute faces and dresses for females, potentially reinforcing soft and fragile stereotypes [20]. The lack of diversity in authoritative avatar designs (e.g., muscular females, effeminate males, gender fluid or neutral designs) might lead to homogenization and further stereotype how authority is asserted and perceived in society. In other words, our avatar designs themselves likely reflect broad gender stereotypes. The diversity in design of avatar robots should be considered so that limited representation does not further reinforce harmful stereotypes [85].

5.3 Can Avatar Robots Mitigate Gender Stereotypes in Authority?

We believe avatar robots have the potential to mitigate gender stereotype problems. More than half of participants perceived authority through behavior observations instead of traditional stereotype. Some customer participants reported that female avatars exhibited stronger behavior than male one. As we discussed in section 4.5.3, the anonymity of avatars could help operators safely exhibit authoritative behaviors by hiding behind the avatar robot. The avatar appeared to reduce social pressure or maintaining a social image and increase safety. Therefore, it could help those who express less authority in their working space to be more assertive, common for women [54, 86]. Interactions like those in our experiments could help with perspective taking training for male operators, like in virtual spaces [12, 32]. Anonymity allowed males to break from modern stereotypes: a male operator (53 years old) enjoyed using a female avatar because the voice resembled a favorite animation character. In this way, male and female participants could act differently from their traditional images, which might blur the borderline between genders. Currently, research on non-binary and gender-neutral robot designs in relation to stereotypes is ongoing [35, 76] and yet to be explored with avatar robots. Studying such avatars across the gender spectrum may help explore or mitigate modern gender stereotype problems in society at large.

However, in contrast to our belief, there are concerns that using avatars might not mitigate the gender stereotypes and even reinforce them. As operators felt stronger self-authority using male avatars, it is possible that people who believe in gender stereotypes will simply replace female avatars with male avatars, which could magnify the stereotype that authoritative tasks are for male avatars. In fact, three participants (2 females, 1 male) voluntarily said if female avatars would be challenged easily, then just use male avatars. However, this encourages gender stereotypes in the long-term. Instead of ignoring the problem and inviting such an unwelcome future, we should study the ways to increase female avatars' expression and perception of authority, perhaps by use in exposure training or enhancing robot itself.

5.4 Does Japanese Culture Play a Role in Perception of Authority?

As Japan is a male-dominated society [47], women have less chance of working in executive roles in companies [3, 11], and receive lower wages [2]. The stronger social image of men may explain why operators experienced a higher self-authority when using a male avatar. In interviews, operators made comments such as "In society, men decide everything" or "In society, people listen to males". Some female operators also shared they experienced discrimination while working as shopkeepers, like how difficult customers tend to be more aggressive towards female shopkeepers than male shopkeepers. Even women would use apologetic strategies to response to sexist abuse [84].

Since Japanese people are more used to a patriarchy, they may have less concern about discussing stereotype-related topics, so that participants openly expressing stereotypical opinions in interviews. For instance, when a male customer tried to guess the actual gender of the operator, he directly said the operator could

not express the rules clearly, similar to a female. If the observations reflect Japanese society, we wonder if people in countries with more awareness and pushback towards gender stereotypes might avoid openly discriminating. Would they treat the gendered avatar robots equally? Would their self-reported behavior match their actions? Cross-cultural research is needed to explore these questions.

6 LIMITATIONS AND FUTURE WORK

As we measured perceived authority with a single Likert item, non-parametric tests were recommended for analysis [75]. However, there is no true alternative non-parametric test to replace the 2 by 2 analysis, so we applied ANOVAs to some of our non-normally distributed single Likert item data. Though previous works showed robustness with Type I error [14, 16, 38, 61, 62, 74] as well as Type II error [30, 38, 43, 74], whether ANOVA is robust for data that violates these assumptions is a topic of debate, especially in HCI work that frequently uses this method [75]. Our analysis method may not address all concerns about statistical methodology. However, as an early work exploring the topic of authority and avatar robots, our results serve as a starting point for future exploration of these types of effects, and can be explored in replications in additional situations. This, when combined with our results, will provide more clarity than any single experiment.

Our role-play method presents limitations, such as potentially unrealistic interactions, though some shopkeeper participants said they experienced similarly difficult customers and stereotypes in real life. Future work may consider new methods to study authority in the real world ethically.

We didn't investigate authority expression in face-to-face human interactions. Many operators used their imagination to compare this with their avatar experiment experience: some participants thought the avatar's authority seemed weaker than a human's due to the lack of physical cues for asserting authority. Our experiment cannot confirm this, and further study is required to compare the differences between avatar robots and real humans.

7 CONCLUSION

We observed how avatar and operator genders influence authority in our role-play experiment with 41 participant pairs, particularly regarding traditional gender stereotypes when avatar robots need to assert authority. Operators felt more authoritative with male avatars. However, the effect on customers' perceptions of a shopkeeper's authority or their time spent on arguing with the shopkeepers were not detected. This suggests avatar robots are still encounter societal gender stereotypes. Less than half the participants attributed their perceptions to stereotypes, while the rest did not. Avatar embodiment, offering physical anonymity, might aid operators in asserting authority. Our findings open pathways for investigating how avatar robots can counter gender stereotypes as they become increasingly prevalent.

ACKNOWLEDGMENTS

This work was supported by JST Moonshot R and D under Grant Number JPMJMS2011, Japan. The authors express gratitude to Hiromi Kobayashi and Yukari Hasegawa for their assistance in the experiments.

REFERENCES

- [1] 2022. *Lawson opens first avatar-staffed convenience store in Tokyo*. Retrieved November 28, 2022 from <https://www.japantimes.co.jp/news/2022/11/28/business/corporate-business/lawson-first-avatar-store/>
- [2] 2022. *OECD Gender wage gap*. Retrieved 2022 from <https://www.oecd.org/tokyo/statistics/gender-wage-gap-japanese-version.htm>
- [3] 2023. *Japan wants women to hold 30 percent of top companies' executive roles by 2030*. Retrieved August 9, 2023 from <https://www3.nhk.or.jp/nhkworld/en/news/backstories/2502/>
- [4] 2023. *Lawson addresses labor shortage with futuristic avatars*. Retrieved August 31, 2023 from <https://www.asahi.com/ajw/articles/14981680>
- [5] Siddharth Agrawal and Mary-Anne Williams. 2017. Robot authority and human obedience: A study of human behaviour using a robot security guard. In *Proceedings of the companion of the 2017 ACM/IEEE international conference on human-robot interaction*. 57–58.
- [6] Sun Joo Ahn, Joshua Bostick, Elise Ogle, Kristine L Nowak, Kara T McGillicuddy, and Jeremy N Bailenson. 2016. Experiencing nature: Embodying animals in immersive virtual environments increases inclusion of nature in self and involvement with nature. *Journal of Computer-Mediated Communication* 21, 6 (2016), 399–419.
- [7] Alexander Mois Aroyo, T Kyohei, Tora Koyama, Hideyuki Takahashi, Francesco Rea, Alessandra Scutti, Yuichiro Yoshikawa, Hiroshi Ishiguro, and Giulio Sandini. 2018. Will people morally crack under the authority of a famous wicked robot?. In *2018 27th IEEE international symposium on robot and human interactive communication (RO-MAN)*. IEEE, 35–42.
- [8] Domna Banakou, Parasuram D Hanumanthu, and Mel Slater. 2016. Virtual embodiment of white people in a black virtual body leads to a sustained reduction in their implicit racial bias. *Frontiers in human neuroscience* (2016), 601.
- [9] John A Bargh, Mark Chen, and Lara Burrows. 1996. Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of personality and social psychology* 71, 2 (1996), 230.
- [10] Marine Beaudoin, Julien Barra, Louise Dupraz, Pauline Mollier-Sabet, and Michel Guerraz. 2020. The impact of embodying an “elderly” body avatar on motor imagery. *Experimental Brain Research* 238, 6 (2020), 1467–1478.
- [11] Melanie Belarmino and Melinda R Roberts. 2019. Japanese gender role expectations and attitudes: A qualitative analysis of gender inequality. *Journal of International Women's Studies* 20, 7 (2019), 272–288.
- [12] Kevin Beltran, Cody Rowland, Nicki Hashemi, Anh Nguyen, Lane Harrison, Sophie Engle, and Beste F Yuksel. 2021. Reducing implicit gender bias using a virtual workplace environment. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–7.
- [13] Jacob T Biehl, Daniel Avrahami, and Anthony Dunnigan. 2015. Not really there: Understanding embodied communication affordances in team perception and participation. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. 1567–1575.
- [14] MJ Blanca, R Alarcón, J Arnao, R Bono, and R Bendayan. 2017. Non-normal data: Is ANOVA still a valid option? *Psicothema*, 29 (4), 552–557.
- [15] Lieselotte Blommaert, Roza Meuleman, Stefan Leenheer, and Anete Butkēviča. 2020. The gender gap in job authority: Do social network resources matter? *Acta Sociologica* 63, 4 (2020), 381–399.
- [16] C Alan Boneau. 1960. The effects of violations of assumptions underlying the t test. *Psychological bulletin* 57, 1 (1960), 49.
- [17] Jorun Børsting and Alma Leora Culén. 2016. A Robot Avatar: Easier Access to Education and Reduction in Isolation? (2016).
- [18] James V Bradley. 1978. Robustness? *Brit. J. Math. Statist. Psych.* 31, 2 (1978), 144–152.
- [19] Sheryl Brahnam and Antonella De Angeli. 2012. Gender affordances of conversational agents. *Interacting with Computers* 24, 3 (2012), 139–153.
- [20] Liudmila Bredikhina and Agnès Giard. 2022. Becoming a Virtual Cutie: Digital Cross-Dressing in Japan. *Convergence* 28, 6 (2022), 1643–1661.
- [21] Noel Brett. 2022. Why Do We Only Get Anime Girl Avatars? Collective White Heteronormative Avatar Design in Live Streams. *Television & New Media* 23, 5 (2022), 451–461.
- [22] Linda L Carli. 1999. Gender, interpersonal power, and social influence. *Journal of social issues* 55, 1 (1999), 81–99.
- [23] Mark Chen and John A Bargh. 1997. Nonconscious behavioral confirmation processes: The self-fulfilling consequences of automatic stereotype activation. *Journal of Experimental Social Psychology* 33, 5 (1997), 541–560.
- [24] Gerald L Clore and Katharine M Jeffery. 1972. Emotional role playing, attitude change, and attraction toward a disabled person. *Journal of personality and social psychology* 23, 1 (1972), 105.
- [25] Jane M Craig and Carolyn W Sherif. 1986. The effectiveness of men and women in problem-solving groups as a function of group gender composition. *Sex Roles* 14 (1986), 453–466.
- [26] Alice H Eagly. 1983. Gender and social influence: A social psychological analysis. *American Psychologist* 38, 9 (1983), 971.
- [27] Alice H Eagly, Christa Nater, David I Miller, Michèle Kaufmann, and Sabine Sczesny. 2020. Gender stereotypes have changed: A cross-temporal meta-analysis of US public opinion polls from 1946 to 2018. *American psychologist* 75, 3 (2020), 301.
- [28] Alice H Eagly and Wendy Wood. 1991. Explaining sex differences in social behavior: A meta-analytic perspective. *Personality and social psychology bulletin* 17, 3 (1991), 306–315.
- [29] Friederike Eysell and Frank Hegel. 2012. (s) he's got the look: Gender stereotyping of robots 1. *Journal of Applied Social Psychology* 42, 9 (2012), 2213–2230.
- [30] Eric B Ferreira, Marcela C Rocha, and Diego B Mequelino. 2012. Monte Carlo evaluation of the ANOVA's F and Kruskal-Wallis tests under binomial distribution. *Sigmae* 1, 1 (2012), 126–139.
- [31] Laura Fiorini, Alessandra Sorrentino, Mattia Pistoletti, Claudia Becchimanzi, Francesca Tosi, and Filippo Cavallo. 2022. Living with a telepresence robot: results from a field-trial. *IEEE Robotics and Automation Letters* 7, 2 (2022), 5405–5412.
- [32] Gili Freedman, Melanie C Green, Max Seidman, and Mary Flanagan. 2021. The Effect of Embodying a Woman Scientist in Virtual Reality on Men's Gender Biases. (2021).
- [33] Cristina Gonzalez-Lienres, Luis E Zapata, Guillermo Iruretagoyena, Sofia Seinfeld, Lorena Perez-Mendez, Jorge Arroyo-Palacios, David Borland, Mel Slater, and Maria V Sanchez-Vives. 2020. Being the victim of intimate partner violence in virtual reality: first-versus third-person perspective. *Frontiers in psychology* 11 (2020), 820.
- [34] Victoria Groom, Jeremy N Bailenson, and Clifford Nass. 2009. The influence of racial embodiment on racial bias in immersive virtual environments. *Social Influence* 4, 3 (2009), 231–248.
- [35] Stefano Guidi, Latisha Boor, Laura van der Bij, Robin Poppen, Okke Rikmenspoel, 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, 615–626.
- [36] Shigeru Haga and Naoki Mizukami. 1996. Japanese version of NASA Task Load Index Sensitivity of its workload score to difficulty of three different laboratory tasks. *The Japanese journal of ergonomics* 32, 2 (1996), 71–79.
- [37] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.
- [38] Michael R Harwell, Elaine N Rubinstein, William S Hayes, and Corley C Olds. 1992. Summarizing Monte Carlo results in methodological research: The one-and two-factor fixed effects ANOVA cases. *Journal of educational statistics* 17, 4 (1992), 315–339.
- [39] Guy Hoffman, Jodi Forlizzi, Shahar Ayal, Aaron Steinfeld, John Antanitis, Guy Hochman, Eric Hochendoner, and Justin Finkenauer. 2015. Robot presence and human honesty: Experimental evidence. In *Proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction*. 181–188.
- [40] Kristina Howansky, Leigh S Wilton, Danielle M Young, Samantha Abrams, and Rebekah Clapham. 2021. (Trans) gender stereotypes and the self: Content and consequences of gender identity stereotypes. *Self and Identity* 20, 4 (2021), 478–495.
- [41] Searle Huh and Dmitri Williams. 2010. Dude looks like a lady: Gender swapping in an online game. *Online worlds: Convergence of the real and the virtual* (2010), 161–174.
- [42] Zaheer Hussain and Mark D Griffiths. 2008. Gender swapping and socializing in cyberspace: An exploratory study. *CyberPsychology & Behavior* 11, 1 (2008), 47–53.
- [43] GK Kanji. 1976. Effect of non-normality on the power in analysis of variance: A simulation study. *International Journal of Mathematical Educational in Science and Technology* 7, 2 (1976), 155–160.
- [44] Teru Kawakita and Hideaki Kanai. 2023. Avatar Design for Bullying Prevention in the Metaverse: Avatar Appearances and the Presumption of Bullying. *Kansei Engineering* (2023), 52.
- [45] Laura J Kray, Leigh Thompson, and Adam Galinsky. 2001. Battle of the sexes: gender stereotype confirmation and reactance in negotiations. *Journal of personality and social psychology* 80, 6 (2001), 942.
- [46] Jong-Eun Roselyn Lee, Clifford I Nass, and Jeremy N Bailenson. 2014. Does the mask govern the mind?: Effects of arbitrary gender representation on quantitative task performance in avatar-represented virtual groups. *Cyberpsychology, Behavior, and Social Networking* 17, 4 (2014), 248–254.
- [47] Jackie FK Lee. 2019. In the pursuit of a gender-equal society: Do Japanese EFL textbooks play a role? *Journal of Gender Studies* 28, 2 (2019), 204–217.
- [48] Min Kyung Lee and Leila Takayama. 2011. “Now, i have a body” uses and social norms for mobile remote presence in the workplace. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 33–42.
- [49] Carol Livingstone. 1983. *Role Play in Language Learning*. ERIC.
- [50] Sarah Lopez, Yi Yang, Kevin Beltran, Soo Jung Kim, Jennifer Cruz Hernandez, Chelsy Simran, Bingkun Yang, and Beste F Yuksel. 2019. Investigating implicit gender bias and embodiment of white males in virtual reality with full body visuomotor synchrony. In *Proceedings of the 2019 CHI Conference on human factors in computing systems*. 1–12.

- [51] Esther Lopez-Zafra and Rocio Garcia-Retamero. 2012. Do gender stereotypes change? The dynamic of gender stereotypes in Spain. *Journal of Gender Studies* 21, 2 (2012), 169–183.
- [52] Risa Maeda, Dražen Brščić, and Takayuki Kanda. 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*. 83–91.
- [53] Paul R Messinger, Xin Ge, Eleni Stroulia, Kelly Lyons, Kristen Smirnov, and Michael Bone. 2008. On the relationship between my avatar and myself. *Journal For Virtual Worlds Research* 1, 2 (2008).
- [54] Gwen Moore and Gene Shackman. 1996. Gender and authority: A cross-national study. *Social Science Quarterly* (1996), 273–288.
- [55] Helen Morgan, Amanda O'donovan, Renita Almeida, Ashleigh Lin, and Yael Perry. 2020. The role of the avatar in gaming for trans and gender diverse young people. *International journal of environmental research and public health* 17, 22 (2020), 8617.
- [56] Daichi Morimoto, Jani Even, and Takayuki Kanda. 2020. Can a robot handle customers with unreasonable complaints?. In *Proceedings of the 2020 ACM/IEEE international conference on human-robot interaction*. 579–587.
- [57] Daphne A Muller, Caro R Van Kessel, and Sam Janssen. 2017. Through Pink and Blue glasses: Designing a dispositional empathy game using gender stereotypes and Virtual Reality. In *Extended abstracts publication of the annual symposium on computer-human interaction in play*. 599–605.
- [58] Clifford Nass and Youngme Moon. 2000. Machines and mindlessness: Social responses to computers. *Journal of social issues* 56, 1 (2000), 81–103.
- [59] Carman Neustaedter, Gina Venolia, Jason Procyk, and Daniel Hawkins. 2016. To Beam or not to Beam: A study of remote telepresence attendance at an academic conference. In *Proceedings of the 19th acm conference on computer-supported cooperative work & social computing*. 418–431.
- [60] Hannah-Hanh D Nguyen and Ann Marie Ryan. 2008. Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of applied psychology* 93, 6 (2008), 1314.
- [61] Geoff Norman. 2010. Likert scales, levels of measurement and the “laws” of statistics. *Advances in health sciences education* 15 (2010), 625–632.
- [62] Egon S Pearson. 1931. The analysis of variance in cases of non-normal variation. *Biometrika* (1931), 114–133.
- [63] Tabitha C Peck, Sofia Seinfeld, Salvatore M Aglioti, and Mel Slater. 2013. Putting yourself in the skin of a black avatar reduces implicit racial bias. *Consciousness and cognition* 22, 3 (2013), 779–787.
- [64] 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*. 428–436.
- [65] Aaron Powers and Sara Kiesler. 2006. The advisor robot: tracing people's mental model from a robot's physical attributes. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*. 218–225.
- [66] Aaron Powers, Adam DI Kramer, Shirlene Lim, Jean Kuo, Sau-lai Lee, and Sara Kiesler. 2005. Eliciting information from people with a gendered humanoid robot. In *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005*. IEEE, 158–163.
- [67] 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 *Social Robotics: 7th International Conference, ICSR 2015, Paris, France, October 26-30, 2015, Proceedings 7*. Springer, 554–563.
- [68] Natalia Reich-Stiebert and Friederike Eyssele. 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*. 166–176.
- [69] David Reitman, Paula C Rhode, Stephen DA Hupp, and Cherie Altobello. 2002. Development and validation of the parental authority questionnaire—revised. *Journal of psychopathology and Behavioral Assessment* 24 (2002), 119–127.
- [70] Zoey Reyes and Joshua Fisher. 2022. The Impacts of Virtual Reality Avatar Creation and Embodiment on Transgender and Genderqueer Individuals in Games: A grounded theory analysis of survey and interview data from Transgender and Genderqueer individuals about their experiences with Avatar Creation Interfaces in Virtual Reality. In *Proceedings of the 17th International Conference on the Foundations of Digital Games*. 1–9.
- [71] Joshua J Reynolds, Victoria Estrada-Reynolds, and Narina Nunez. 2018. Development and validation of the Attitudes Towards Police Legitimacy Scale. *Law and Human Behavior* 42, 2 (2018), 119.
- [72] Laurie A Rudman and Stephen E Kilianski. 2000. Implicit and explicit attitudes toward female authority. *Personality and social psychology bulletin* 26, 11 (2000), 1315–1328.
- [73] Toni Schmader, Michael Johns, and Chad Forbes. 2008. An integrated process model of stereotype threat effects on performance. *Psychological review* 115, 2 (2008), 336.
- [74] Emanuel Schmider, Matthias Ziegler, Erik Danay, Luzi Beyer, and Markus Bühner. 2010. Is it really robust? *Methodology* (2010).
- [75] Mariah Schrum, Muyleng Ghuy, Erin Hedlund-Botti, Manisha Natarajan, Michael Johnson, and Matthew Gombolay. 2023. Concerning trends in likert scale usage in human-robot interaction: Towards improving best practices. *ACM Transactions on Human-Robot Interaction* 12, 3 (2023), 1–32.
- [76] Katie Seaborn and Peter Pennefather. 2022. Neither “hear” nor “their”: Interrogating gender neutrality in robots. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 1030–1034.
- [77] Brett Sherrick, Jennifer Hoewe, and T Franklin Waddell. 2014. The role of stereotypical beliefs in gender-based activation of the Proteus effect. *Computers in Human Behavior* 38 (2014), 17–24.
- [78] Aaron Steinfeld, Terrence Fong, David Kaber, Michael Lewis, Jean Scholtz, Alan Schultz, and Michael Goodrich. 2006. Common metrics for human-robot interaction. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*. 33–40.
- [79] Kazuaki Takeuchi, Yoichi Yamazaki, and Kentaro Yoshifuji. 2020. Avatar work: Telework for disabled people unable to go outside by using avatar robots. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. 53–60.
- [80] Masami Tamagawa. 2018. Coming out of the closet in Japan: an exploratory sociological study. *Journal of GLBT Family Studies* 14, 5 (2018), 488–518.
- [81] Benedict Tay, Younbo Jung, and Taezoon Park. 2014. When stereotypes meet robots: the double-edge sword of robot gender and personality in human-robot interaction. *Computers in Human Behavior* 38 (2014), 75–84.
- [82] Benedict Tiong Chee Tay, Taezoon Park, Younbo Jung, Yeow Kee Tan, and Alvin Hong Yee Wong. 2013. When stereotypes meet robots: The effect of gender stereotypes on people's acceptance of a security robot. In *Engineering Psychology and Cognitive Ergonomics. Understanding Human Cognition: 10th International Conference, EPCE 2013, Held as Part of HCI International 2013, Las Vegas, NV, USA, July 21-26, 2013, Proceedings, Part I 10*. Springer, 261–270.
- [83] Brandon Van Der Heide, Erin M Schumaker, Ashley M Peterson, and Elizabeth B Jones. 2013. The Proteus effect in dyadic communication: Examining the effect of avatar appearance in computer-mediated dyadic interaction. *Communication Research* 40, 6 (2013), 838–860.
- [84] Katie Winkle, Ryan Blake Jackson, Gaspar Isaac Melsión, Dražen Brščić, Iolanda Leite, and Tom Williams. 2022. Norm-breaking responses to sexist abuse: A cross-cultural human robot interaction study. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 120–129.
- [85] Katie Winkle, Donald McMillan, Maria Arnelid, Katherine Harrison, Madeline Balaam, Ericka Johnson, and Iolanda Leite. 2023. Feminist human-robot interaction: Disentangling power, principles and practice for better, more ethical HRI. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*. 72–82.
- [86] Erik Olin Wright, Janeen Baxter, and Gunn Elisabeth Birkelund. 1995. The gender gap in workplace authority: A cross-national study. *American sociological review* (1995), 407–435.
- [87] Meir Yaish and Haya Stier. 2009. Gender inequality in job authority: A cross-national comparison of 26 countries. *Work and Occupations* 36, 4 (2009), 343–366.
- [88] Nick Yee and Jeremy Bailenson. 2007. The Proteus effect: The effect of transformed self-representation on behavior. *Human communication research* 33, 3 (2007), 271–290.
- [89] Atsushi Yoshida, Hirokazu Kumazaki, Taro Muramatsu, Yuichiro Yoshikawa, Hiroshi Ishiguro, and Masaru Mimura. 2022. Intervention with a humanoid robot avatar for individuals with social anxiety disorders comorbid with autism spectrum disorders. *Asian Journal of Psychiatry* 78 (2022), 103315.