Interacting with Computers, 2026, 1–17
https://doi.org/10.1093/iwc/iwag030
Published: 8 June 2026
Long article

OXFORD

The design of neutrality: evaluations and biases in the perception of male, female, and neutral avatar faces in a gender diverse sample

Lorenzo Reina *, Stefano Guidi, Oronzo Parlangeli

Department of Social, Political and Cognitive Sciences (DISPOC), University of Siena, Siena, Italy

*Corresponding author: Via Orione, 9, Cervia, Italy. Email: lorenzo.reina@unisi.it.

Abstract

Gender stereotypes can influence user interactions with digital agents, shaping perceptions and behaviors based on avatar representation. However, gender-neutral designs have only recently emerged as a potential strategy to challenge and reduce these biases. Moreover, the effectiveness of such designs remains largely underexplored in the context of non-binary individuals, who are often excluded from evaluation studies due to their low representation. In an online survey experiment using a 3x3 mixed factorial design and a validated set of 9 stimuli, participants of different self-reported genders evaluated masculine, feminine, and gender-neutral avatar faces across multiple perceptual dimensions (likeability, agency, communion, interaction, and attractiveness), together with measures of ambivalent sexism (ASI) and ambivalence toward men (AMI). The results revealed a strong effect of participant gender on avatars perception, with non-binary participants providing significantly lower ratings than both men and women for all dependent variables. When gender-prejudice measures were included, these group differences were substantially reduced, revealing that dimensions of ambivalent sexism and ambivalence toward men shaped avatar perception. Benevolence toward men increased ratings only among male participants, whereas hostility toward men enhanced evaluations across all participants. Hostile sexism also interacted with both participant and avatar gender, influencing agency and communion judgments. Overall, the results indicate that rather than pursuing the search for seemingly gender-neutral technologies, it is more appropriate to include in the design processes of these technologies, participants with different genders and different attitudes toward the multiplicities of gender expression, and to employ more gender-diverse rather than gender-neutral avatar representations.

RESEARCH HIGHLIGHTS

- Using a 3×3 mixed factorial online experiment, 359 participants identifying as men, women, and non-binary evaluated nine computer-generated avatar faces (masculine, feminine, and gender-neutral) across multiple perceptual dimensions including likeability, agency, communion, and willingness to interact.
- Avatar gender alone did not significantly influence stereotypical attributions or social perceptions, suggesting that subtle facial cues are insufficient to activate gender-based expectations in human–agent interaction.
- Non-binary participants rated all avatar types significantly lower than men and women across all perceptual dimensions, revealing a pronounced negative bias toward the avatars regardless of their gender presentation.
- Individual gender prejudice measures, including hostile and benevolent sexism and ambivalence toward men, substantially accounted for group differences in avatar perception, highlighting the central role of attitudinal factors in human–agent interaction.
- Gender neutrality in avatar design does not emerge as a universally inclusive solution; gender-diverse, expressive, and customizable representations are recommended to better serve users across the full spectrum of gender identities.

Keywords digital avatars, gender-neutral, gender bias, non-binary, human–agent interaction

Received: March 6, 2025. **Revised:** March 13, 2026. **Accepted:** March 23, 2026

Introduction

Gender stereotypes permeate various aspects of human life, including how we design and interact with technology—especially as anthropomorphic digital agents become increasingly prevalent (Beckwith & Burnett, 2004; Bryson et al., 2020; O'Connor & Liu, 2023). These systems, which replicate human traits in both form and behavior, often rely on culturally embedded cues, including gender, to elicit familiarity and trust. However, these same cues risk reinforcing narrow social expectations: gender biases—belief systems that prescribe normative definitions of what it means to be a man or a woman—are frequently embedded, both implicitly and explicitly, within interactive technologies (Golombok & Fivush, 1994; Spiel, 2021).

While gender has long been a subject of analysis in social sciences such as historiography, gender studies, and queer theory, where it is understood as fluid and multifaceted (Bornstein, 1994; Brooks, 2017; Butler, 1990, 1999; Norton, 2009), this complexity is rarely mirrored in technological systems. To navigate the complex landscape of gender identities, we adopt the perspective of Thorne et al. (2019), who emphasize how terminology in this field is constantly evolving and deeply linked to lived and personal experience. In this work, the term "non-binary" is used broadly to describe all identities that fall outside the male/female dichotomy, including those who perceive their gender as fluid or completely absent (Broussard et al., 2018). Within this framework, we use the expression "gender-neutral" to refer to those agents' identities that actively avoid binary marking, seeking a space that is neutral with respect to cultural norms (Seaborn & Pennefather, 2022b). We also use the term "gender-ambiguous" to describe those manifestations that mix or alternate masculine and feminine traits in ways that make the identity not immediately categorizable (Sutton, 2020).

The human capacity to navigate this broad and situationally modulated spectrum of gender expression (Beischel et al., 2022; Joel et al., 2015) contrasts sharply with the binary logic often imposed by technical infrastructures, which tend to marginalize or erase gender-nonconforming people (Broussard, 2018; Hamidi et al., 2018). As a result, non-binary and gender-diverse individuals are underrepresented not only in the design of such systems, but also in their empirical evaluation—an exclusion that contributes to the reproduction of discriminatory assumptions in human–robot interaction (HRI) (Amiri et al., 2024; Seaborn et al., 2023; Winkle et al., 2023a).

In recent years, scholars in HRI have critically examined the implications of gendering artificial agents, highlighting how even minimal cues such as voice, role, or appearance can activate gender stereotypical expectations (Eyssel & Hegel, 2012; Tay et al., 2014; Torre et al., 2023). Consequently, growing attention has also been devoted to the design of gender-neutral or gender-ambiguous agents, with the aim of reducing bias and creating more inclusive systems (Danielescu, 2020; De Cet et al., 2025; Parlangeli et al., 2023b). However, it still appears difficult to determine what induces the perception of gender neutrality, and furthermore, this appears to be deeply context-dependent. Some scholars argue that neutral design, if not informed by participatory and inclusive methods, may default to normative assumptions—often reproducing masculine-coded traits or failing to resonate with non-binary users (McDonnell & Baxter, 2019; Seaborn & Pennefather, 2022b).

Despite this growing body of research, current experimental studies often focus on binary gender conditions and predominantly recruit cisgender participants, leaving open questions about how neutral agents are perceived by individuals who themselves identify outside the gender binary (Atwood et al., 2024; Stolp-Smith & Williams, 2024). As such, a key limitation of this field lies in the underrepresentation of non-binary voices—not only as users, but as epistemic stakeholders in the evaluation of design.

This study responds directly to this call for inclusion by explicitly recruiting non-binary participants and analyzing their responses in comparison with cisgender groups. This approach not only increases the ecological validity of our findings but also aligns with ongoing efforts in HRI to develop technologies that reflect the diversity of the people who interact with them. By focusing on user evaluations of stimuli that are perceived as gender-neutral, feminine or masculine, and by including participants of diverse gender identities, we aim to assess whether neutral design supports more balanced perceptions and reduces the salience of gendered associations in human–agent interaction. Specifically, the experiment was conducted using computer-generated avatar faces previously validated as feminine, masculine, or gender-neutral, and involving samples of women, men, and non-binary participants. The experiment used a 3 × 3 mixed factorial design, with repeated measures on avatar gender; each participant evaluated three avatars per gender category. Individual measures of gender stereotypes were also included.

The paper is organized as follows: Section "Related Works" reviews prior research on gendered design in HRI and outlines ongoing debates surrounding gender-neutral and ambiguous agents. Section "Experiment" describes the experiment, detailing the methods (experimental design, including materials, sample, and measures), and presenting and discussing the results. Section "Conclusions" draws conclusions from the overall results, with attention to both perceptual and attitudinal differences among participants of varying gender identities.

Related works

Gender and stereotypes in human–agent interaction

Research consistently demonstrates that users attribute gender to artificial agents in response to minimal social cues, including voice characteristics, physical appearance, and functional roles (Aylett et al., 2023; Bernotat et al., 2021; Eyssel & Perugia, 2024; Perugia et al., 2022). This is because intelligent agents, including humanoid and social robots, virtual conversational agents, and chatbots, function as genuine social actors in human interactions (Nass & Moon, 2000; Reeves & Nass, 1996). This social actor paradigm emerges from two key phenomena: first, individuals maintain consistent interpersonal communication patterns when engaging with these agents (Craiut & Iancu, 2022); second, it reflects our fundamental biological predisposition to interpret experiences through the lens of human relationships (Fogg, 1998; Norman, 2007).

Within this framework of social categorization, the gendering of robots has emerged as a particularly significant area of investigation (Alesich & Rigby, 2017; Parlangeli et al., 2023a; Søraa, 2017; Torre et al., 2023). The pervasiveness of this gendering process has led numerous researchers to systematically investigate how robot gender will be manipulated and perceived across various experimental settings, revealing the relevance of integrating multiple sensory and contextual cues, including appearance, voice, role, and behavior (Dennler et al., 2025; Perugia & Lisy, 2023; Tay et al., 2014), in alignment with social stereotypes (Eyssel & Hegel, 2012; Robertson, 2010).

Among these various cues, voice has emerged as a particularly influential factor in gender perception, and the increasing use of voice assistants has led to growing interest in gender perception in synthetic voices (Law et al., 2021a; Raghunath et al., 2022; Seaborn et al., 2021; Tolmeijer et al., 2021; West et al., 2019). Moreover, as demonstrated by a recent study, the perceived gender of voice assistants influences user trust and reliance (Zhang et al., 2025). Assistants perceived as female are rated as more trustworthy compared to male ones, particularly in contexts involving critical health information, such as medication instructions (Goodman & Mayhorn, 2023).

However, this effect appears to be moderated by the shape of the robot. Steinhaeusser et al. (2021) found that manipulating the robot's voice gender (male, female, or neutral) did not significantly impact user attitudes, anthropomorphism, or narrative engagement. The authors suggest that the low human-likeness of the NAO robot used in the experiment may have prevented users from forming gendered expectations in the first place—indicating that voice cues alone may be insufficient without visual or behavioral congruence.

In other cases, a robot's gender is manipulated to correlate with the gender stereotype of its operational role.

For example, female robots are often designed as companions for the elderly, babysitters, and cleaning assistants, as well as sexual robots, while those with masculine presentations typically receive assignments involving authority, technical expertise, or security operations (Bryant et al., 2020; Richardson, 2016). For instance, a study conducted with 198 individuals in Singapore revealed that participants felt more comfortable with the idea of a robot performing security-related tasks if it looked masculine, and with a robot performing household cleaning tasks if it looked feminine (Tay et al., 2014). Furthermore, users are more likely to share personal information with robots that have male voices, suggesting that gender stereotypes subtly but powerfully shape not only the perceived credibility of robots, but also other aspects of the social interaction they elicit (Behrens et al., 2018). In some cases, the congruence between a robot's perceived gender and the gender stereotypically referred to a task increases perceived competence (Law et al., 2021b). However, Bryant et al. (2020) caution that such congruence does not necessarily improve user trust, and may instead entrench bias by validating stereotype-based role assignments.

Similarly, gender perception in robots is often shaped by the congruence between their visual presentation and the stereotypical expectations associated with their functional role. Visual gender manipulations—such as variations in facial features, hairstyle, accessories, and body proportions—have been widely employed to shape the perceived gender of robots (Bernotat et al., 2021; Eyssel & Hegel, 2012; Neuteboom & de Graaf, 2021; Paetzel et al., 2016; Parlangeli et al., 2023b). These visual cues are not interpreted in isolation but interact with contextual expectations about the robot's role. For instance, Nomura and Suzuki (2022) found that users with stronger gender stereotypical beliefs were more likely to associate female-appearing robots with roles like caregiving or customer service. However, individuals professionally involved in those roles, such as nurses or counselors, showed less reliance on gendered expectations—suggesting that domain-specific expertise may reduce the influence of visual gender cues.

These findings are consistent with additional experimental evidence showing how manipulations of appearance and behavioral features interact with user expectations and gender biases. For instance, Ye et al. (2020) demonstrated that both the gender and anthropomorphic design of a robot influenced how users responded to a service failure in a simulated retail setting. Participants tended to penalize male robots more severely for errors compared to female robots, particularly when the robot appeared more human-like. Interestingly, male users preferred machine-like female robots, while female users favored human-like male robots, suggesting that user gender also shaped perceptions of likeability and humanness. These results underscore how the intersection of anthropomorphism and gender activates social expectations that directly affect trust and forgiveness.

This pattern was further confirmed by Galatolo et al. (2023), who examined robots challenging gender stereotypes. In their study with 400 participants, humanoid robots with either male or female presentation discussed women's underrepresentation in Science, Technology, Engineering, and Mathematics (STEM) or men's low parental-leave uptake. Their findings revealed that perceived credibility was influenced by both the robot's gender presentation and the stereotype being addressed—male robots were generally rated as more credible, particularly when addressing stereotypes about men, while female robots were perceived as less credible when countering stereotypes about women. This effect was especially pronounced among male observers.

As shown above, gendering in HRI illustrates how social stereotypes shape human–agent interactions. At the same time, these patterns also reflect deeper dynamics of exclusion and power embedded in design practices. Such dynamics become especially evident when artificial agents reproduce normative assumptions without accounting for the diversity of users and contexts, resulting in unintended yet systematic forms of harm.

While designers often leverage gender attributions to promote familiarity or acceptance, the practice of "gendering" robots carries the risk of perpetuating harmful and limiting social norms (Bryant et al., 2020; West et al., 2019; Winkle et al., 2021). This is not incidental: design is inherently political, as it reflects the values and assumptions of its creators (Winner, 2017). In the context of robotics, these embedded values can reproduce stereotypes and disproportionately harm marginalized groups—particularly when those communities are excluded from the design process (Costanza-Chock, 2018). This issue extends beyond gender, affecting people with intersecting marginalized identities. For example, UNESCO has reported that the predominance of men in the development of voice assistants has led to the creation of submissive, abuse-tolerant female personas, reinforcing gendered stereotypes (Weisman & Janardhanan, 2020). Similarly, the lack of diversity among AI developers has contributed to biased systems that perpetuate discriminatory narratives across race, gender, and ability (Hundt et al., 2025; Birhane et al., 2021). Large language models have been shown to produce harmful outputs, including sexist, racist, and ableist content (Gross, 2023; Hartvigsen et al., 2022), while perception systems often misclassify or exclude non-binary and multiply marginalized individuals (Buolamwini & Gebru, 2018; Scheuerman et al., 2019).

To counteract these harms, scholars have proposed ethical design frameworks rooted in *Participatory Design* and *Design Justice* (Costanza-Chock, 2020; Spinuzzi, 2005). These approaches call for a redistribution of design agency by involving marginalized users in shaping robotic systems. For instance, the Feminist HRI Framework (Winkle et al., 2023b), the Equitable Design Framework (Ostrowski et al., 2022), and Robots for Social Justice (Zhu et al., 2024) all promote methods to center user experiences, expose power imbalances, and mitigate systemic bias in HRI.

The present investigation situates itself within this theoretical framework by critically examining gender perception in human–agent interaction while exploring design strategies that transcend binary representations. Through the deliberate inclusion of gender-neutral avatars and the consideration of non-binary users' perspectives, this study advances the scholarly discourse toward a more inclusive approach to HRI—one that acknowledges the socio-political dimensions of technological design and endeavors to challenge normative assumptions embedded in human–machine interactions.

Gender-neutral and ambiguous agents: between potential and perception

Gender neutrality, in its simplest and most basic form, refers to the absence of gender and/or the lack of cues that will lead to the attribution of a specific gender (Udry, 1994). Another way of defining gender neutrality within human–computer interaction (HCI) and HRI is through gender fluidity, which refers to a shift in gender identity or expression over time (Katz-Wise, 2020). Fluidity involves envisioning robots or artificial agents designed to embrace multiple gender representations simultaneously, with interchangeable and overlapping components and, above all, the potential to work towards normalizing non-binary gender representations while familiarizing people with non-traditional ideas about gender (Dudek & Young, 2022).

The concept of gender neutrality in artificial agents has gained traction in HRI as a design strategy aimed at reducing the reinforcement of binary gender norms and associated stereotypes—although some researchers argue that achieving true gender neutrality may be technically and socially unattainable (Seaborn & Pennefather, 2022a; Seaborn & Pennefather, 2022b; Torre et al., 2023).

In response to these challenges, several research initiatives have explored the use of ambiguous or androgynous design features. One of the most significant projects addressing gender neutrality in the field of artificial intelligence and digital assistants is Project Q. Copenhagen Pride, Virtue, Equal AI, Koalition Interactive, and Thirty Sounds Good collaborated to create a genderless voice aimed at ending gender biases in AI-based voice assistants (Danielescu, 2020). The design team developed a vocal frequency positioned at the intersection of male voice ranges, typically between 85 and 180 Hz, and female ranges, generally between 140 and 255 Hz.

In another project, Parlangeli et al. (2023b) created nine computer-generated avatar faces, which 30 participants evaluated for gender and age through an online questionnaire. From the results of these tests, three faces were selected—one male, one female, and one neutral—and further assessed through an online questionnaire regarding certain feminine or masculine personality traits and the willingness/pleasure to interact with artificial agents possessing those faces. The findings revealed that the stimulus characterized by greater gender neutrality was less likely to be associated with a feminine stereotype (Parlangeli et al., 2023b).

The success of these strategies depends on a combination of factors, such as the alignment of a neutral voice with other gender cues and with observers' expectations, with prior work showing that ambiguous voices can reduce the tendency to ascribe a gender to robots both in neutral contexts (De Cet et al., 2025) and even when the robot displays stereotypically gendered attributes (Torre et al., 2023). However, in the end, the result of these efforts may appear unclear and unnatural, confirming that fully removing gender-coded cues from voices remains a complex challenge (Yu et al., 2022).

Yet, neutrality does not guarantee inclusivity. Seaborn and Pennefather (2022a, 2022b), through a critical review, argue that so-called neutral agents often inherit dominant norms, implicitly aligning with masculine defaults or erasing non-binary experiences. Their analysis distinguishes between "gender neutrality"—the absence of gender cues—and "gender ambiguity"—a blending of cues that may still allow for identity recognition without forcing categorization. Similarly, McDonnell and Baxter (2019) caution that while gender-neutral chatbots may be seen as more professional or unbiased, users may also perceive them as impersonal or emotionally distant. Likewise, Danielescu et al. (2023) argue that while non-binary users prefer voice assistants that reflect their identity, mainstream users may initially respond to these voices as "strange" or "unintelligent"—a response shaped more by cultural conditioning than by actual system performance.

Despite these tensions, the literature increasingly suggests that ambiguous or neutral design—particularly in voice—will be a viable tool for disrupting binary categorization and reducing stereotype activation.

Including the non-binary perspectives

Although HRI has long acknowledged the importance of gender in robot design, most research in this field continues to lack explicit documentation of the participation of transgender and non-binary individuals. According to several scholars (Perugia & Lisy, 2023; Seaborn et al., 2023; Seaborn & Frank, 2022; Seaborn & Pennefather, 2022b; Winkle et al., 2023a), there is a striking absence of non-cisgender individuals in HRI studies.

However, research has begun to address this gap. In a recent study, Atwood et al. (2024) investigated the potential influence of individuals' awareness of their gender category on the process of perceiving gender in faces. The study examined three groups of participants—gender diverse (trans, non-binary, genderqueer), cisgender-LGBPQ+, and cisgender-heterosexual—and found that gender diverse participants were significantly more accurate in classifying more androgynous faces compared to both cisgender groups. This suggests that long-term life experiences and continual exposure to androgynous faces are factors that should not be overlooked in perception analysis (Atwood et al., 2024). Stolp-Smith and Williams (2024) conducted interviews with transgender and non-binary users, finding that many viewed robot design as an opportunity to subvert traditional gender norms, emphasizing fluidity, pronoun choice, and inclusive language as essential design concerns.

Additional efforts have sought to include non-binary communities in the development of technological systems, particularly voice assistants. Hope's research focuses on developing gender-inclusive synthetic voices and voice technologies based on data from non-binary and transgender communities (Hope & Székely, 2025; Székely & Hope, 2024), promoting inclusive design for voice assistants and language-based robots.

The previously mentioned Project Q also included non-binary individuals in its evaluation of a gender-neutral voice assistant. The results showed that 4,600 non-binary evaluators from Denmark, the United Kingdom, and Venezuela were evenly split between those who reported being unable to distinguish the voice's gender and those who perceived it as either male or female (Carpenter, 2019). Building on this participatory approach, Danielescu et al. (2023) developed "Sam," a non-binary TTS voice co-designed with input from non-binary and transgender individuals. Their large-scale study revealed that non-binary users were significantly more likely than cisgender users to prefer a non-binary voice, particularly when it exhibited a balanced combination of masculine and feminine vocal markers. This project exemplifies how participatory processes can inform inclusive voice design and reinforces the importance of centering marginalized perspectives in the development of speech-based technologies.

Experiment

Research questions and hypotheses

The experiment is conducted to investigate three sets of research questions about the perception of digital avatars representing virtual assistants, focusing on the perception of gender-neutral avatars in comparison to feminine and masculine avatars, among both non-binary individuals and individuals who identify with a binary gender (men and women). For all sets of research questions we tested several hypotheses derived from previous research in social psychology, HCI, and HRI.

RQ1. Does the gender of a digital avatar influence: (1) people's perception of the avatar's stereotypical traits such as agency and communion (Gray & Wegner, 2012; Hsu et al., 2021), (2) people's willingness to interact with the avatar, and (3) the avatar's likeability?

First, in light of existing gender stereotypes we expect that gender-neutral avatars will be less subject to gender stereotypical attributions, which consider men as higher in agency than women, and women higher in communion than men, leading to the following predictions:

- (H1.1a) feminine but not neutral avatars will be attributed less agency than masculine avatars;
- (H1.1b) masculine but not neutral avatars will be attributed less communion than feminine avatars.

In accordance with previous findings about the effects of gendering robots (Behrens et al., 2018; Bryant et al., 2020; Law et al., 2021b; Tay et al., 2014) we also expect that:

- (H1.2) feminine avatars will be rated more positively than masculine avatars, concerning dimensions such as interaction, likeability, and attractiveness.

It is less clear how the neutral avatar will be perceived concerning these aspects in comparison with feminine and masculine avatars. Moreover, it is plausible that the perception of different avatars could vary across participants of different gender identities, which leads us to the second set of questions.

RQ2. Do individuals of different (self-identified) genders differ in their perceptions of avatars discernible as having a different gender, concerning: (1) the avatar's stereotypical personality traits, (2) their willingness to interact with the avatar, and (3) the avatar's likeability?

Research in social psychology and HRI brings us to make several predictions concerning the interplay between participant gender and avatar gender. First, social categorization of the avatar in terms of gender can lead to the outgroup-homogeneity effect (Heitlinger et al., 2022; Park & Rothbart, 1982). We can then expect that:

- (H2.1a) attributions of personality traits to feminine avatars will be more stereotypical in men than in women and attributions to masculine avatars will be more stereotypical in women than in men.

In light of the fact that gender stereotypes reflect and support power and status differences between men and women as social groups (Bareket & Fiske, 2023; Eagly & Koenig, 2021; Fiske et al., 2002) to the advantage of men (the majority group), we can also expect that:

- (H2.1b) non-binary individuals, being a minority group, will be less prone to stereotyping by gender and will react to gendered avatars in a way that is more similar to women than to men.

Social categorization can also lead to the in-group bias (Fraune et al., 2020 ; Tajfel, 1971; Tajfel & Turner, 1979; Verkuyten, 2021), leading us to formulate a series of hypotheses:

- (H2.2a) men will view masculine avatars more favorably than feminine ones, and evaluate them more positively than women do;
- (H2.2b) women will evaluate feminine avatars more favorably than masculine ones, and more positively than men do;
- (H2.2c) non-binary participants will evaluate neutral avatars more positively than other groups and more positively than masculine and feminine avatars. This effect will be strengthened by the appreciation by this population of gender representations in digital technologies that do not reflect a binary conception (Atwood et al., 2024; Danielescu et al., 2023; Morgenroth & Ryan, 2018).

In addition, we tested further hypotheses related to the role of sexism and gender prejudice in the perception of gendered avatars in women, men, and non-binary persons.

RQ3. Do the individual levels of prejudice against men and women modulate the effect of the gender of an avatar on: (1) people's perception of the avatar's stereotypical traits, (2) people's willingness to interact with the avatar, and (3) the avatar's likeability?

We expect that higher levels of hostile or benevolent sexism, as measured by the Ambivalent Sexism Inventory (ASI) subscales, and higher levels of *ambivalence* toward men, as measured by the Ambivalence toward Men Inventory (AMI) subscales (Glick & Fiske, 1996, 1999; Rollero et al., 2014), will be associated with stronger stereotypical attributions. For example, participants endorsing benevolent sexist beliefs are expected to attribute higher communion and lower agency scores to feminine avatars, while those high in hostile sexism will attribute higher agency and lower communion scores to them. This leads to the following predictions:

- (H3.1) Higher levels of hostile or benevolent sexism (ASI) and ambivalence toward men (AMI) will be associated with stronger stereotypical attributions.

We also expect that participants' willingness to interact with the avatars will be influenced by their individual levels of gender prejudice. In particular, we expect that:

- (H3.2a) higher levels of hostility toward women will be associated with less positive views of feminine avatars;
- (H3.2b) higher levels of hostility toward men will be associated with less positive views of masculine avatars.

Method

Design and participants

The design of the experiment was a 3 × 3 mixed factorial design, with replication. Participants of different self-reported genders (men, women, and non-binary) evaluated avatars belonging to three gender categories (feminine, masculine, and neutral), with each category represented by three different faces validated in previous work (Parlangeli et al., 2023b) (see Section "Stimuli").

We aimed to recruit 360 English-speaking participants from the United States, the United Kingdom, and Ireland—120 for each of the three gender groups. Participants were recruited through the Prolific platform. A power analysis conducted with G-power showed that the sample size would have 95% power to detect a small interaction between avatar gender and participant gender. In preparation of participants' exclusion for missing inclusion criteria, we initially recruited 381 participants across the three gender groups. Following data quality checks and consistency screening between self-declared gender identity and responses provided in the survey, 22 participants were excluded. This resulted in a final sample of 359 participants, which was used for all subsequent analyses. The sample included individuals self-identifying as women, men, or non-binary, equally distributed across conditions (men = 120, women = 119, non-binary = 120). The presentation order of the nine avatars (three per avatar gender) was randomized for each participant. Completion time was approximately 15 min. Each participant received 1 GBP (≈ 1.17 €) for their participation in the study. This research was approved by the ethical committee of the University of Siena (Careus prot. 0223954; 20, 11, 25).

Procedure and measures

The experiment was implemented as an online survey using the LimeSurvey platform. Initially participants were informed about the nature of the task, the procedure, and their rights. They were told that we were developing a virtual assistant and that we wanted to collect feedback on a prototype. After participants had expressed their consent to participate, they were presented with nine avatar faces, one at a time, and for each one they had to rate their perception of the avatar on the same semantic differential and Likert-scale items assessing likeability, agency, communion, willingness to interact and attractiveness (see Table 1).

Stereotypes

To test the activation of gender stereotypes, we measured the perception of the avatars along the two fundamental dimensions of social and mind perception—agency and communion—using 10 items based on the short form of the Bem Sex Role Inventory (Choi et al., 2009). Five questions formed a coherent scale to evaluate *agency* (α = .93), a trait typically associated with masculinity in social stereotypes. Another five questions assessed *communion* (α = .93), a stereotypically feminine trait. Responses were expressed on a 7-point Likert scale (1 = *Absolutely not*; 7 = *Absolutely yes*), and averaged across items for each dimension.

Interaction

The dimension of interaction was constructed using five items validated by the study of Parlangeli et al. (2023b) and adapted from items in scales from various studies (Huang et al., 2022; Scutella et al., 2022; Venkatesh & Davis, 2000). Participants were asked to express their willingness to interact with the avatar whose face they saw on a 7-point Likert scale (1 = Absolutely not; 7 = Absolutely yes). Responses were averaged across items to yield a composite measure, with excellent internal consistency (α = .91). Nonetheless, we decided to exclude from the scale the items related to beauty and realism, as we reasoned that the former was more related to emotional responses than to interaction, while the latter addressed a different facet of artificial agents perception—namely, anthropomorphism. The exclusion of these two items further increased the internal consistency of the scale (α = .94).

Emotional responses

Emotional responses were investigated along the dimension of likeability. Drawn from studies by Monahan (1998) and Bartneck et al. (2009), the dimension of likeability included five items on a 5-point semantic differential scale (α = .92).

Gender prejudice

In addition to these measures, we included a section comprising the short versions of the ASI and the AMI (Glick & Fiske, 1996, 1999; Rollero et al., 2014). The ASI, which assesses hostile and benevolent attitudes toward women (12 items, 6 for each subscale, α_{ASI} = .92, α_{HS} = .93, α_{BS} = .87), was particularly relevant for the evaluation of the feminine-looking avatar; the AMI, assessing ambivalent attitudes toward men (12 items, 6 for each subscale, α_{AMI} = .9, α_{HM} = .8, α_{BM} = .92), was included as it could influence the perception of the masculine-looking avatar.

Demographics, gender, and gender conformity

Finally, the questionnaire included demographic questions about age (years), and gender, measured both with a multiple-choice question offering the options male, female, non-binary, and I'd rather not say, and using the bipolar version of the Traditional Masculinity Femininity Scale (TMFS), developed by Kachel et al. (2016) to measure conformity to traditional gender norms along the bipolar masculine–feminine continuum. This scale includes six items measuring self-conformity with traditional gender definitions (self-identification, attitudes, behavior, and appearance), each answered on a 7-point scale from 1 = completely masculine to 7 = completely feminine.

Stimuli

The images of the avatar faces used in this experiment were derived from a set of stimuli validated in previous studies for gender differentiation. Specifically, the avatar faces employed were evaluated in a study by Parlangeli et al. (2023b). This research tested nine computer-generated faces (Parlangeli et al., 2023b), including three feminine (F1, F2, F3), three masculine (M1, M2, M3), and three neutral avatars (N1, N2, N3) (see Table 2), and identified significant differences across the four dimensions of masculinity, femininity, gender neutrality, and age.

Based on these results, three representative faces were selected: M1 for the male gender, F2 for the female gender, and N2 for the neutral gender. These stimuli were deemed suitable for conducting future experiments (Parlangeli et al., 2023b) (Figure 1).

Statistical analyses

To analyze the data, we conducted a series of linear mixed-effects models (LLMs) using R (lme4 package), with participant ID as a random effect to account for repeated measures across the three avatar faces.

Table 1 Summary of the questionnaire structure, including each measured dimension, corresponding items, literature sources, and scale formats.

Dimension	Question/item	References	Scale type
Likeability	Dislike–like; unfriendly–friendly; unkind–kind; unpleasant–pleasant; awful–nice	Monahan (1998); Bartneck et al. (2009)	5-point semantic differential
Agency	To what extent do you think this avatar could: –behave as a leader? –advocate their own ideas? –have a strong personality? –make important decisions? –stand their own ground?	Choi et al. (2009)	7-point Likert scale (1 = absolutely not; 7 = absolutely yes)
Communion	To what extent do you think this avatar could: –show affection? –feel empathy? –be tender? –show compassion? –be kind?	Choi et al. (2009)	7-point Likert scale (1 = absolutely not; 7 = absolutely yes)
Interaction	–To what extent would you trust this avatar? –would you like to interact with this avatar? –would you like this avatar to help you while studying or working?	Parlangeli et al. (2023b); Huang et al. (2022); Venkatesh and Davis (2000)	7-point Likert scale (1 = absolutely not; 7 = absolutely yes)
Attractiveness	–To what extent do you find this avatar attractive?	Parlangeli et al. (2023b)	Slider (visual analog scale) from 0 ("not at all") to 100 ("very much")

Table 2 Descriptive statistics for the sample in the study by participants' self-reported gender.

Men N = 120	**Non-binary N = 120**		**Women N = 119**	
Variable	***M* (*SD*)**	***M* (*SD*)**	***M* (*SD*)**	***p*-value**
Age	32.7^a (6.67)	29.3^b (6.19)	32.7^a (7.09)	$<$.001
Gender identity (TMFS)	2.07^a (1.38)	4.03^b (0.90)	5.81^c (1.08)	$<$.001
Hostile sexism (HS)	2.82^a (1.16)	1.12^b (1.01)	2.76^a (1.17)	$<$.001
Benevolent sexism (BS)	2.30^a (1.42)	0.46^b (0.90)	1.90^c (1.33)	$<$.001
Hostility toward men (HM)	2.84^a (1.43)	0.67^b (0.85)	2.44^c (1.36)	$<$.001
Benevolence toward men (BM)	2.12^a (1.10)	1.67^b (1.12)	2.54^c (1.05)	$<$.001

Note. Within each row, means that do not share a superscript (*a*, *b* or *c*) differ ($p < .05$). BS = benevolent sexism; BM = benevolence toward men; HS = hostile sexism; HM = hostility toward men; TMFS = traditional masculinity-femininity scale.

Fixed effects included participant gender, avatar gender category, and their interaction.

We adopted a multilevel structure to account for both participant-level and stimulus-level random variability, directly addressing the concern of stimulus sampling bias (Wells & Windschitl, 1999) and accounting for individual variability in response to conditions. Significance was evaluated at α = .05, and family-wise error was controlled using the Holm-Bonferroni correction where applicable across multiple dependent variables. All the data and code for the analyses are available on OSF.

Results

Sample characteristics

Participants' mean age was 31.6 (SD = 6.83). We computed descriptive statistics (means and *SDs*) across the three groups of participants for all covariates (Table 3) and compared the means across groups using ANOVA and pairwise comparisons. Non-binary participants were significantly younger and self-reported a significantly queer sexual orientation than the other groups. Concerning conformity to traditional gender norms (on a masculine to feminine continuum), non-binary participants were self-reporting as significantly more feminine than men and significantly less feminine than women. Non-binary participants also had significantly lower scores than the other groups on all the measures of gender prejudice.

Hypotheses testing

The results of the LMMs including both participant-level and stimulus-level random variability (Table 4), revealed for all the dependent variables a significant main effect of participant gender (all $p < .001$). No other significant main effects or interactions were found.

Concerning the main effect of participant gender, pairwise comparisons (Table 5) revealed that for all dependent variables non-binary participants gave significantly lower ratings than both men and women, which in turn were not significantly different from each other in the mean ratings for avatar perception.

Effects of gender prejudice

To examine the effect of the gender prejudice on the avatars perception, we fitted further LMMs in which we included as predictors of the ratings also the participants' scores on the ASI and AMI subscales (*hostile sexism*, *benevolent sexism*, *hostility toward men*, and *benevolence toward men*), in interaction with both participant gender group and avatar gender. All the covariates were centered on the mean before

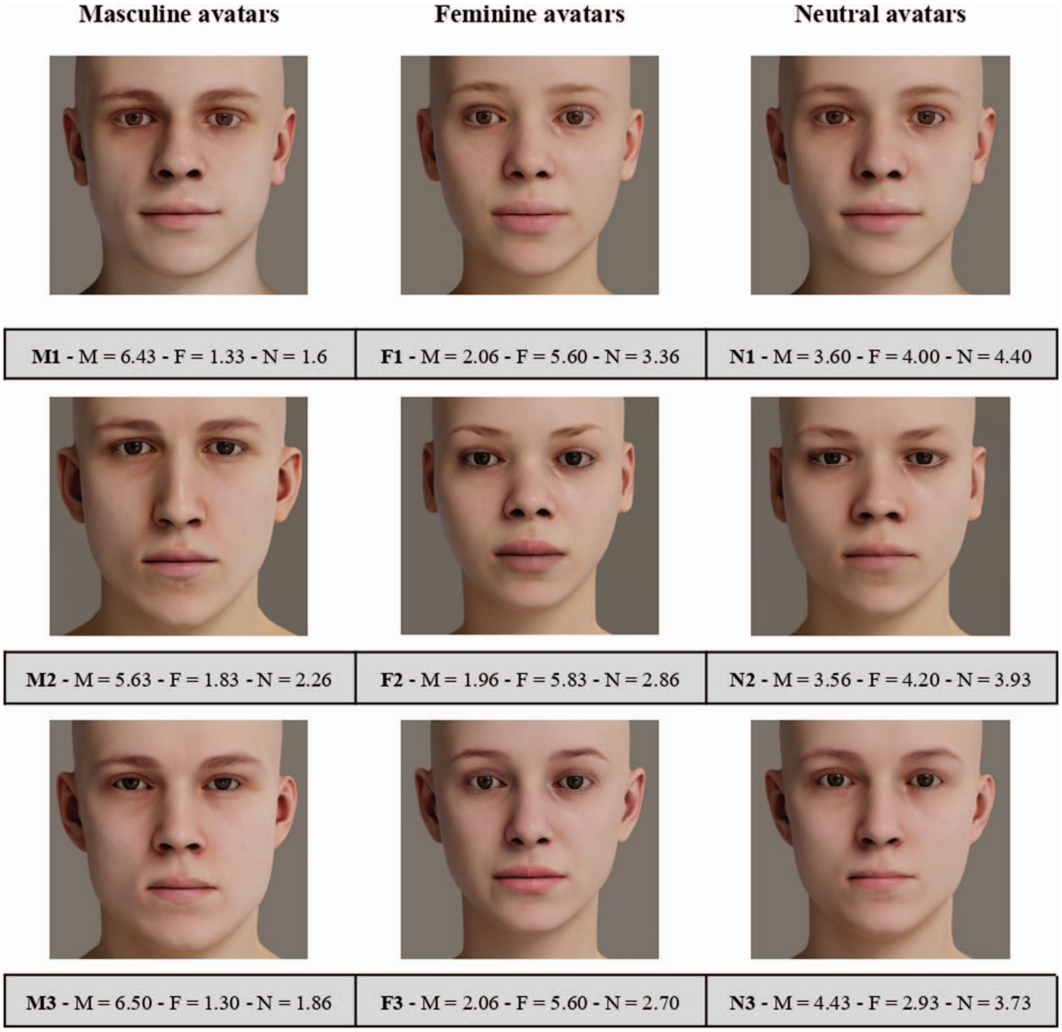

Figure 1 Stimuli used in the study for the masculine (left column), feminine (center) and neutral (right) avatars. For each avatar, we report the mean scores for masculinity, femininity, and neutrality obtained for these avatars in a previous validation study (Parlangeli et al., 2023b).

Table 3 Results of the text of the main effects and interaction in the LMMs for all the dependent variables in the study.

	Avatar gender		**Participant gender**		**AG × PG**	
DV	***F***	***p***	***F***	***p***	***F***	***p***
Agency	1.00	.421	42.99**	.001	0.83	.628
Communion	1.49	.421	33.47**	.001	1.35	.417
Interaction	1.25	.421	58.00**	.001	0.65	.628
Likeability	1.16	.421	27.96**	.001	2.52	.155
Attractiveness	2.40	.421	40.00**	.001	2.24	.155

Note. For each effect tested, p values were adjusted for a family of five dependent variables, using the false discovery rate method.
AG = avatar gender; DV = dependent variable; LMM = linear mixed model; PG = participant gender; * p< .05, ** $p < .01$, *** $p < .001$.

fitting the models. The results of the tests of the main effects and interactions for all the variables are reported in Table 5, while all the parameters of the LMMs (including estimated predictors coefficients for the fixed effects and random variances and correlations for random effects) are reported as supplementary materials on OSF.

The results (Table 5) showed that with the covariates the main effect of participant gender was only significant for interaction (p = .025) and attractiveness (p = .025), while it was no longer significant for agency, communion and likeability. Among the covariates, first of all *Order* had a small but significant positive effect on all the dependent variables (all $p < .001$). *Hostility toward men* (HM) was also significantly and positively (B_{AG} = 0.27, B_{CO} = 0.25, B_{IN} = 0.30, B_{LI} = 0.13, B_{AT} = 6.57) associated with all the dependent variables (p = .015–.044), and interacted significantly with avatar gender for attractiveness only (p = .030).

Benevolence toward men (BM) interacted significantly with participant gender for all the dependent variables (p = .004). Finally, the results showed a significant three-way interaction between avatar

Table 4 Marginal means and *SEs* for the ratings of the avatar perception dimensions in the experiment as a function of participant gender.

	Women		Non-binary		Men	
DV	***M***	***(SE)***	***M***	***(SE)***	***M***	***(SE)***
Agency	4.41[a]	(0.13)	3.26[b]	(0.13)	4.52[a]	(0.13)
Communion	4.16[a]	(0.2)	3.15[b]	(0.2)	4.36[a]	(0.2)
Interaction	4.20[a]	(0.19)	2.77[b]	(0.19)	4.52[a]	(0.19)
Likeability	3.35[a]	(0.14)	2.73[b]	(0.14)	3.38[a]	(0.14)
Attractiveness	49.69[a]	(2.69)	28.09[b]	(2.68)	52.76[a]	(2.68)

Note. For each dimension (row) the means that do not share the same superscript (*a*, *b* or *c*) are significantly different at the $p < .05$. DV = dependent variable.

Table 5 Results of the text of the main effects and interaction in the linear mixed-effects models, including the covariates for all the dependent variables in the study.

Effect	**Agency**		**Communion**		**Interaction**		**Likeability**		**Attractiveness**	
	F	***p***	***F***	***p***	***F***	***p***	***F***	***p***	***F***	***p***
Avatar gender	0.70	.530	1.44	.530	1.12	.530	0.95	.530	1.99	.530
Participant gender	1.31	.340	0.96	.383	**4.69***	**.025**	1.69	.310	**5.15***	**.025**
HM	**8.35***	**.015**	**6.53***	**.018**	**7.74***	**.015**	**4.10***	**.044**	**5.13***	**.030**
BM	0.10	.938	0.01	.938	0.31	.938	0.23	.938	0.88	.938
HS	2.31	.325	2.97	.325	0.60	.549	0.95	.549	0.07	.799
BS	0.00	.974	0.17	.974	1.53	.974	0.05	.974	0.69	.974
Order	**45.74****	**.001**	**39.83****	**.001**	**52.38****	**.001**	**52.18****	**.001**	**104.55****	**.001**
AG × PG	0.46	.924	0.60	.924	0.23	.924	0.47	.924	0.47	.924
AG × HM	0.55	.957	1.19	.765	0.03	.970	0.09	.970	**5.10***	**.030**
PG × HM	2.41	.152	2.96	.140	1.82	.205	2.90	.140	0.87	.418
AG × BM	0.85	.688	0.60	.688	0.98	.688	0.81	.688	0.15	.860
PG × BM	**6.03****	**.004**	**5.54****	**.004**	**5.90****	**.004**	**6.17****	**.004**	**11.01****	**.004**
AG × HS	1.39	.413	0.11	.892	3.18	.210	0.96	.479	1.97	.350
PG × HS	0.24	.933	0.07	.933	0.18	.933	0.18	.933	0.31	.933
AG × BS	0.36	.698	0.94	.488	1.14	.488	1.46	.488	1.70	.488
PG × BS	0.21	.972	0.53	.972	0.28	.972	0.03	.972	0.56	.972
AG × PG × HM	0.73	.581	1.44	.545	0.72	.581	1.10	.581	1.63	.545
AG × PG × BM	0.72	.605	1.58	.440	0.68	.605	2.59	.175	0.78	.605
AG × PG × HS	**2.96***	**.048**	**2.95***	**.048**	1.06	.486	1.03	.486	0.51	.729
AG × PG × BS	0.59	.983	0.10	.983	0.29	.983	0.16	.983	0.94	.983

Note. For each effect tested, *p* values were adjusted for a family of five dependent variables, using the FDR method. AG = avatar gender; BM = benevolence toward men; HS = hostile sexism; HM = hostility toward men; Order = order of avatar presentation; PG = participant gender; * $p < .05$, ** $p < .01$, *** $p < .001$;. Boldface indicate statistically significant effects.

gender (AG), participant gender (PG), and *hostile sexisms* (HS) for both agency (p = .048) and communion (p = .048).

We followed up the significant interactions using trend analysis and interaction contrasts. For each dependent variable and covariate, we report only the analyses of the highest order interactions.

Concerning the interaction between *participant gender* and *benevolence toward men*, as shown in Figure 2A–E, the results showed that in men benevolence toward men tended to increase ratings for all the dependent variables, while it tended to decrease ratings in women and non-binary participants in a very similar way. Interaction contrasts for the trends confirmed that the slope for the effect of BM was significantly greater (and positive) for men than for the other groups combined (p = <.001–.001) for all the variables, while the slopes for women and non-binary participants were always negative and not significantly different from each other (p = .178–.806).

Additionally, the positive effect of *hostility toward men* was greater for the masculine avatar (B = 9.63) than for the other avatars combined (p = .002), while the slopes for the neutral (B = 5.89) and feminine avatars (B = 4.20) were not significantly different from each other, p = .332). The slopes, however, were positive for all the avatars, but they were only significantly different from zero for the masculine (p = .002) and neutral avatars (p = .03).

We analyzed the three way interactions between *hostile sexism, participant gender,* and *avatar gender* found on agency and communion ratings (Figure 3A–B), by computing for each participant group two interaction contrasts: one (a) between the slope for the feminine avatar and the slopes for the masculine and neutral avatars (combined), and one (b) between the slope for the masculine and the slope for the neutral avatar. For both dependent variables, the first contrast was significant for women (p_{Agency} = .008, p_{Com} = .013) and non-binary (p_{Agency} = .046, p_{Com} = .024) participants, and for agency only also marginally for men (p_{Agency} = .064, p_{Com} = .254), while the second contrast was never significant regardless of the group ($p > .20$).

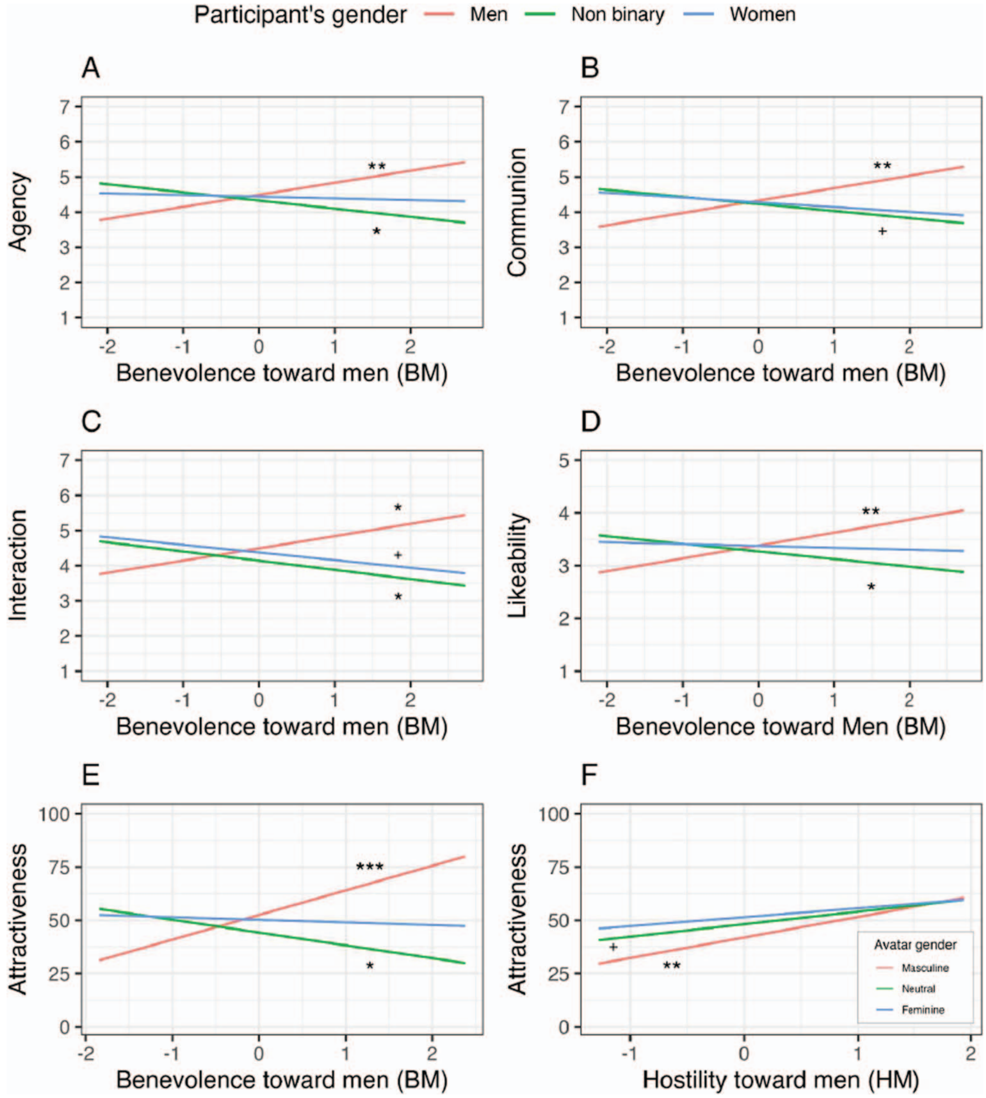

Figure 2 (A–E) Plots of the effect of benevolence toward men on the ratings for all the robot-perception variables by participant gender. (F) Plot of the effect of hostility toward men on attractiveness ratings by avatar gender. Symbols next to the slopes represent the significance levels for the tests of the individual slopes: + $p < .1$, *$p < .05$, **$p < .01$, ***$p < .001$.

Discussion

Lack of influence of avatar gender on social perception

The experiment was conducted to test several hypotheses about the perception of digital avatars representing virtual assistants, concerning the effect of the gender of the avatar on several dimensions of perception (H1.1-2), the way this effect depends on the individual (self-reported) gender of the perceiver (H2.1-2), and the individual level of gender prejudice (HP3.1-3).

We first expected that the gender of the avatar could activate gender stereotypes (Fiske et al., 2002), and thus that feminine avatars will be attributed less agency (H1.1a) and more communion (H1.1b) than masculine avatars, and that gender-neutral avatars could be less affected by stereotypes and would not be considered different from the other avatars in agency and communion. We also expected (H1.2) that feminine avatars could be viewed more favorably than masculine avatars (Behrens et al., 2018; Bryant et al., 2020; Law et al., 2021b; Tay et al., 2014). These predictions were not supported by the results, which did not find a significant effect of avatar gender on any avatar perception variable. It is possible that subtle facial cues alone such as those we manipulated and used to gender the avatars may be insufficient to trigger gender-based perceptions. In this experiment, gender was manipulated using subtle cues and without the stereotypical cues that are often used for this aim (hair, makeup, body shape, clothes). Particularly, the lack of feminine stereotypical cues and of cues related to gender dimorphism concerning biological secondary sexual characteristics (e.g., breasts in women) the gender category could have prevented an activation of female gender stereotypes in participants.

Absence of ingroup bias and interaction between participant and avatar gender

In the light of psychosocial research, we also expected that men and women will show opposite biases depending on the gender of the avatar, with a tendency to (H2.1a) stereotype less (Heitlinger et al., 2022; Park & Rothbart, 1982) and (H2.2a-b) judge more favorably

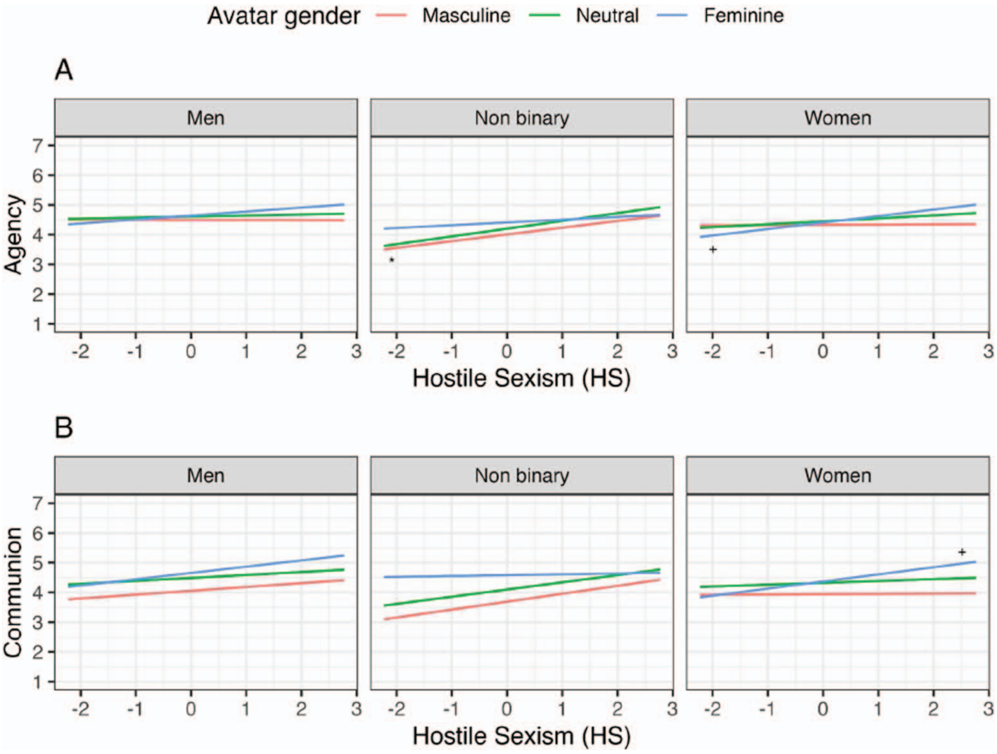

Figure 3 (A–B) Plots of the effect of benevolence toward men on the ratings for (A) agency and (B) communion by avatar and participant gender. Symbols next to the slopes represent the significance levels for the tests of the individual slopes: $+\ p < .1$, $^{*}p < .05$, $^{**}p < .01$, ***, $p < .001$.

the avatar of their own gender (ingroup) than the avatar of the opposite gender (outgroup) (Fraune et al., 2020 ; Tajfel, 1971; Tajfel & Turner, 1979; Verkuyten, 2021), with each group judging the same-gender avatar more favorably than the other group. We also hypothesized that non-binary individuals could judge the neutral avatar (ingroup) more favorably than the other groups and more favorably than the feminine and masculine avatars (H2.2c) (Atwood et al., 2024; Danielescu et al., 2023; Morgenroth & Ryan, 2018). None of these hypotheses was supported by the results of the experiment, in which the interaction between participant gender and avatar gender was not significant for any of the dependent variables. It is possible that this could again be due to insufficient gender cues in the stimuli, which prevented the activation of gender-related information (Perugia & Lisy, 2023).

The role of gender prejudice and the default male bias

The results showed that sexism variables were associated with avatar perception in various ways, which in some cases depended on participant gender. In general, however, our prediction that individual levels of gender prejudice will predict stereotypical attributions to gendered avatars (H3.1), and modulate their acceptance and emotional reactions (H3.2a/b), was not supported. With a few exceptions, in fact, sexism variables did not interact significantly with avatar gender. However, our results highlighted a general positive association between hostility toward men and avatar ratings, regardless of avatar gender. It is possible that a general positive response bias could be related to this finding, and that participants who tended to give higher ratings for all the avatar perception scales were also more inclined to give positive responses to all the questions included in the sexism scales, which were rated on an agreement scale. A positive effect of hostility toward men was expected for agency attributed to masculine avatar, and the "halo" effect, often found when ratings about social perception dimensions are collected about specific individuals rather than about social categories or groups, could explain the effect on communion ratings and possibly on the other dimensions of avatar perception. The reason this effect was found regardless of avatar gender seems to indicate a default male gender attribution by participants to our avatars, in the light of absence in the stimuli of more salient gender cues. This is consistent with previous findings about gender perception in humanoid robots (Bernotat et al., 2021; Eyssel & Hegel, 2012).

Moreover, we found that benevolence toward men tended to have a significant positive effect on men's ratings, a significant negative effect on non-binary participants' ratings, and a non-significant negative effect for women. This form of gender prejudice is characteristic of a view of men as competent and having high status, and of women as incomplete without a man, toward whom they have the duty to provide support and care (Barreto & Doyle, 2023; Barreto & Ellemers, 2005). So, if the default gender of the avatars was male, we can expect that those higher in this variable will have perceived the (implicitly masculine) avatars more favorably and attributed them more agency, and possibly also more communion. The fact that this positive relation was only found in men seems to suggest that avatar gender will have been perceived differently by men than by the other participants. If men were perceiving the avatars (regardless of their gender) as more masculine than the other groups did, it makes sense that more positive ratings will be attributed by participants showing more benevolence toward men.

Variations of hostile sexism effects across gender groups

We then found that the effect of hostile sexism on stereotypical traits varied both across participant gender and across avatar gender. For non-binary participants the effect was mainly evident and similarly positive for masculine and neutral avatars, but was not found for feminine avatars. For women, the effect of hostile sexisms was still positive, but was only found for feminine avatars, and not for masculine and neutral avatars. For men, lastly, the effect was never significant. Given that this construct refers to a form of gender prejudice which sees women—especially if they do not conform to traditional gender norms—as seeking to take control and power from men's hand (Becker, 2010), we will have expected, especially in men (Bareket & Fiske, 2023), a negative effect of this variable on the perception of feminine avatars. As suggested by Bareket and Fiske (2023), men as the dominant group may be more inclined to use stereotypes to maintain gender-based power and status differences (see also Fiske, 1993), but it is possible that, again, the lack of more explicit gender cues could explain the lack of this effect in men. The fact that for women we found that hostile sexism increased the attribution of agency and communion to feminine avatars, but not to the other avatars, reflects the open questions in the literature regarding how and whether the link between hostile and benevolent sexism and stereotypes is stronger in men than in women (Bareket & Fiske, 2023). We must notice, however, that this effect was small and only marginally significant. For non-binary participants, lastly, the results showed that the effect was basically null for the ratings about feminine avatars, and it was significantly more positive for the other avatars, although the slopes were not significantly different from zero.

Lower avatar ratings among non-binary participants

The results of the experiment, lastly, revealed a strong, unexpected effect of participant gender on avatar perception, with non-binary participants providing significantly lower ratings than both men and women for all the dependent variables. There will be multiple, non-mutually exclusive reasons for this, some related to gender prejudice, and some to attitudes toward gender expression and identity.

First of all, the lower ratings by non-binary participants will be due to an explicit rejection of gender stereotypes. Compared to the other groups, in fact, non-binary participants had also significantly lower scores on all the individual prejudice measures, with larger differences with respect to those typically found between men and women (Rollero et al., 2014). This suggests that participants' ratings will be an expression of a conscious refusal of traditional gender norms and of a binary conceptualization of gender identity, and of the related prescriptive and descriptive gender stereotypes (Morgenroth & Ryan, 2018). Ascribing group differences in avatar perception to differences in the levels of sexism across the groups is consistent with our finding that when the gender prejudice subscales were inserted in the analyses as covariates, the effect of participant gender on avatar perception was either no longer significant (for the stereotypical dimensions and likeability ratings) or its size was reduced.

The sexism scales were administered at the end of the questionnaire, after participants had rated all the avatars, and thus when they could have realized that the gender of the avatar was one of the factors we were interested in. This will have prompted non-binary participants to react by distancing themselves from stereotypes about gender.

But even just the questions used to measure agency and communion—which were the first they had to answer about each avatar—could have made participants realize that the experiment was about gender perception, and prompted their reaction against traditional gender conceptions and prejudice. Indeed, also when supplementary analyses we considered only the responses to the first trial, non-binary ratings for the avatars were significantly lower than men's and women's ratings. But the difference became non-significant again after controlling for individual differences in prejudice. It is worth noting that voluntarily suppressing stereotypical judgments requires cognitive effort (Devine, 1989), and thus we will expect it to be more efficient in the earlier experimental trials than in the later ones, and this is consistent with the positive effect of order found in the experiment. However, if the effect of order was due to a reduction in stereotype suppression due to fatigue, we will expect to find it only, or mainly, in non-binary participants, while it was found similarly in all the groups.

A last further explanation for the differences observed across gender groups is that non-binary participants could have disliked more than the other groups, a representation of gender in the avatars along a bipolar continuum (from masculine to feminine through gender neutrality). This will have been seen as reinforcing a binary conception of gender, which is very different from the complex gender identity of individuals whose self-identify outside traditional gender categories (Morgenroth & Ryan, 2018; Schudson & Morgenroth, 2022). This hypothesis is clearly related to the hypothesis about the role of gender prejudice, as it is likely that lower levels of gender prejudice in non-binary participants could be due to a similar rejection of binarism and gender stereotypes. Moreover, the lack of gender expression features (gender performance) in the avatars will have also been disliked by non-binary participants, whose identity will correspond more to a combination of masculine and feminine characteristics (e.g., a more diverse gender performance and expression) rather than to a lack of gendered characteristics.

Limitations

This study presents several limitations that should be addressed in future research. First, gender expression in the stimuli was conveyed solely through static and subtle facial features, with the avatars cropped at the forehead and devoid of hair, body, voice, or behavior. This narrow representation limits ecological validity, and future studies should consider using avatars with a richer and more nuanced combination of masculine and feminine traits, potentially in immersive virtual reality (VR) environments. Second, the within-subjects design may not have been optimal; by making the manipulation too transparent, it could have amplified participants' awareness of the study's aims and triggered stronger negative reactions, particularly among non-binary participants. A between-subjects design, with repeated exposures to avatars of the same gender, could improve statistical power and better test whether non-binary participants are reacting specifically against gender-stereotyped cues—an effect that will not occur when they are exposed to gender-diverse or androgynous avatars. Additionally, the exclusive use of static stimuli limits the expression of gender through dynamic and socially salient cues such as voice, gesture, clothing, and behavioral style. Future work should explore the impact of these features in more ecologically valid, immersive contexts. Lastly, all measures in the current study were explicit; however, implicit processes may still be influenced by avatar gender. Employing implicit tasks involving facial stimuli could help uncover unconscious biases or distancing behaviors, offering a more comprehensive understanding of how different gender groups respond to virtual agents. Future studies should explore how gender representation, context, individual preferences, and personal experiences shape user perceptions, contributing to meaningful progress in understanding the effects of gender stereotypes and in designing truly accessible and inclusive technologies.

Conclusions

Previous research has analyzed interactions with humanoid robots and artificial agents based on their assigned gender revealing preferences and stereotypes (Bernotat et al., 2021; Eyssel & Hegel, 2012; Fogg, 1998; Guidi et al., 2022; Norman, 2007; Perugia & Lisy, 2023; Tay et al., 2014). Within the context of gender bias studies, gender neutrality in robots and virtual agents should aim at reducing prejudices (Danielescu, 2020). However, transgender and non-binary perspectives are often overlooked in HRI research, despite the importance of including diverse viewpoints (Seaborn et al., 2023; Stolp-Smith & Williams, 2024).

This study sought to expand our understanding of the role of gender in avatar faces by exploring how users perceive male, female, and neutral avatars. The results indicate that gender in virtual agents has little effect on influencing the perception of the avatar and in activating gender stereotypes.

Our main finding is the negative bias expressed by non-binary participants toward the avatars. This result suggests different implications for HCI research and practice. First of all, they imply that researchers should actually avoid gender neutrality and stimuli in which gender characteristics are provided only with minimal facial cues. Although simplified stimuli will provide some control over extraneous variables, they risk activating negative attitudes in non-binary individuals, which in turn could invalidate research and negatively affect the acceptability of and interaction with virtual agents. At the same time, the lack of a preference for neutral faces among non-binary participants and their negative bias suggests that gender neutrality, rather than being a viable, universally inclusive solution for virtual agents, may be actually detrimental. However, our hypothesis about the role of gender-expressive features in non-binary perception could also suggest a way to contrast the negative bias in non-binary participants against gender-neutral avatar. Using selected gender-expressive features in a mixed (and customizable) combination will be perceived by non-binary individuals as a more correct representation of their identity, one that denies binarism. And this, in turn, will not only prevent negative biases, but also foster acceptance and improve interaction with virtual agents.

Overall, our findings enhance our knowledge of the role of gender in digital agent faces and ultimately underscore the importance of adaptive design. Designers should consider the diversity of gender-related preferences and integrate customizable elements, allowing users to select or modify the appearance of avatars according to their needs.

Gender neutrality will not be a solution for overcoming gender stereotypes, but understanding this, and understanding the reasons behind it, could represent an important step toward more inclusive technology.

Data availability

All data and code for the analyses presented in this paper are available in a public OSF repository at the following link: https://osf.io/dvfk5/

Funding

This work was supported by research funds (PSR) from the Department of Social, Political and Cognitive Sciences (DISPOC) of the University of Siena.

Conflicts of interest

The authors declare no conflict of interest.

Acknowledgements

We thank the anonymous reviewers for their valuable feedback.

References

Alesich, S., & Rigby, M. (2017). Gendered robots: Implications for our humanoid future. *IEEE Technology and Society Magazine*, **36**, 50–59. https://doi.org/10.1109/MTS.2017.2696598

Amiri, G. A., Shahidi, S., Darmel, F. A., Mehri, M., & Anwari, M. A. (2024). Beyond binary gender: Navigating gender ambiguity in voice user interfaces. *Cognizance Journal of Multidisciplinary Studies*, **4**, 79–86. https://doi.org/10.47760/cognizance.2024.v04i06.008

Atwood, S., Gibson, D. J., Briones Ramírez, S., & Olson, K. R. (2024). Flexibility in continuous judgments of gender/sex and race. *Journal of Experimental Psychology: General*, **153**, 2931–2950. https://doi.org/10.1037/xge0001512

Aylett, M. P., Carmantini, A., Pidcock, C. J., Nichols, E., Gomez, R., & Siskind, S. R. (2023, March). Haru he's here to help! A demonstration of implementing comedic rapid turn-taking for a social robot. In *Companion of the 2023 ACM/IEEE International Conference on Human–Robot Interaction* (pp. 912–913). https://doi.org/https://dl.acm.org/doi/10.1145/3568294.3580217

Bareket, O., & Fiske, S. T. (2023). A systematic review of the ambivalent sexism literature: Hostile sexism protects men's power; benevolent sexism guards traditional gender roles. *Psychological Bulletin*, **149**, 637–698. https://doi.org/10.1037/bul0000400

Barreto, M., & Doyle, D. M. (2023). Benevolent and hostile sexism in a shifting global context. *Nature Reviews Psychology*, **2**, 98–111. https://doi.org/10.1038/s44159-022-00136-x.

Barreto, M., & Ellemers, N. (2005). The burden of benevolent sexism: How it contributes to the maintenance of gender inequalities. *European Journal of Social Psychology*, **35**, 633–642. https://doi.org/10.1002/ejsp.270.

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, **1**, 71–81. https://doi.org/10.1007/s12369-008-0001-3.

Becker, J. C. (2010). Why do women endorse hostile and benevolent sexism? The role of salient female subtypes and internalization of sexist contents. *Sex Roles*, **62**, 453–467. https://doi.org/10.1007/s11199-009-9707-4.

Beckwith, L., & Burnett, M. M. (2004). Gender: An important factor in end-user programming environments? In *In 2004 IEEE Symposium* *on Visual Languages-Human Centric Computing* (pp. 107–114). IEEE. https://doi.org/10.1109/VLHCC.2004.28.

Behrens, S. I., Egsvang, A. K. K., Hansen, M., & Møllegård-Schroll, A. M. (2018). Gendered robot voices and their influence on trust. In *Companion of the 2018 ACM/IEEE International Conference on Human–Robot Interaction* (pp. 63–64). https://doi.org/10.1145/3173386.3177009.

Beischel, W. J., Schudson, Z. C., Hoskin, R. A., & van Anders, S. M. (2022). The gender/sex 3 × 3: Measuring and categorizing gender/sex beyond binaries. *Psychology of Sexual Orientation and Gender Diversity*, **10**, 355–372. https://doi.org/10.1037/sgd0000558.

Bernotat, J., Eyssel, F., & Sachse, J. (2021). The (fe)male robot: How robot body shape impacts first impressions and trust towards robots. *International Journal of Social Robotics*, **13**, 477–489. https://doi.org/10.1007/s12369-019-00562-7.

Birhane, A., Prabhu, V. U., & Kahembwe, E. (2021). Multimodal datasets: Misogyny, pornography, and malignant stereotypes, arXiv, arXiv:2110.01963. https://doi.org/10.48550/arXiv.2110.01963.

Bornstein, K. (1994). *Gender outlaw: On men, women, and the rest of us*. Routledge.

Brooks, J. (2017, April 24). *Boy? Girl? Both? Neither? A new generation overthrows gender*. KQED. https://www.kqed.org/futureofyou/335790/boy-girl-both-neither-a-new-generation-overthrows-gender

Broussard, K. A., Warner, R. H., & Pope, A. R. (2018). Too many boxes, or not enough? Preferences for how we ask about gender in cisgender, LGB, and gender-diverse samples. *Sex Roles*, **78**, 606–624. https://doi.org/10.1007/s11199-017-0823-2.

Broussard, M. (2018). *Artificial unintelligence: How computers misunderstand the world*. The MIT Press.

Bryant, D. A., Borenstein, J., & Howard, A. (2020, March). Why should we gender? The effect of robot gendering and occupational stereotypes on human trust and perceived competency. In *Proceedings of the 2020 ACM/IEEE International Conference on Human–Robot Interaction* (pp. 13–21). Association for Computing Machinery. https://doi.org/10.1145/3319502.3374778

Bryson, J., Etlinger, S., Keyes, O., Leonard, A., & Rankin, J. L (2020). *Gender bias in technology: How far have we come and what comes next?*. Centre for International Governance Innovation. https://www.cigionline.org/articles/gender-bias-technology-how-far-have-we-come-and-what-comes-next

Buolamwini, J., & Gebru T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In S. A. Friedler, & C. Wilson (Eds.), *Proceedings of the 1st Conference on Fairness, Accountability and Transparency* (pp. 77–91). PMLR. https://proceedings.mlr.press/v81/buolamwini18a.html

Butler, J. (1990). *Gender trouble: Feminism and the subversion of identity*. Routledge.

Butler, J. (1999). *Gender trouble: Feminism and the subversion of identity* (2nd). Routledge.

Carpenter, J. (2019). Why project Q is more than the world's first non-binary voice for technology. *Interactions*, **26**, 56–59. https://doi.org/10.1145/3358912

Choi, N., Fuqua, D. R., & Newman, J. L. (2009). Exploratory and confirmatory studies of the structure of the Bem Sex Role Inventory short form with two divergent samples. *Educational and Psychological Measurement*, **69**, 696–705. https://doi.org/10.1177/0013164409332218

Costanza-Chock, S. (2018). Design justice: Towards an intersectional feminist framework for design theory and practice. In C. Storni, K. Leahy, M. McMahon, P. Lloyd, & E. Bohemia (Eds.), *Design as a catalyst for change – DRS International Conference 2018*. (pp. 25–28). https://doi.org/10.21606/drs.2018.679

Costanza-Chock, S. (2020). *Design justice: Community-led practices to build the worlds we need*. The MIT Press.

Craiut, M. V., & Iancu, I. R. (2022). Is technology gender neutral? A systematic literature review on gender stereotypes attached to artificial intelligence. *Human Technology*, **18**, 297–315. https://doi.org/10.14254/1795-6889.2022.18-3.6

Danielescu, A. (2020, July). Eschewing gender stereotypes in voice assistants to promote inclusion. In *Proceedings of the 2nd Conference on Conversational User Interfaces* (pp. 1–3). Association for Computing Machinery. https://doi.org/10.1145/3405755.3406151

Danielescu, A., Horowit-Hendler, S. A., Pabst, A., Stewart, K. M., Gallo, E. M., & Aylett, M. P. (2023, April). Creating inclusive voices for the 21st century: A non-binary text-to-speech for conversational assistants. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1–17). Association for Computing Machinery. https://doi.org/10.1145/3544548.3581281

De Cet, M., Sturdee, M., Obaid, M., & Torre, I. (2025, March). Sketching robots: Exploring the influence of gender-ambiguous voices on robot perception. In *2025 20th ACM/IEEE International Conference on Human–Robot Interaction (HRI)* (pp. 103–112). IEEE. https://doi.org/10.1109/HRI61500.2025.10973931

Dennler, N., Kian, M., Nikolaidis, S., & Matarić, M. (2025). Designing robot identity: The role of voice, clothing, and task on robot gender perception. *International Journal of Social Robotics*, **17**, 707–728. https://doi.org/10.1007/s12369-025-01209-6

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, **56**, 5–18. https://doi.org/10.1037/0022-3514.56.1.5

Dudek, S. Y., & Young, J. E. (2022, March). Fluid sex robots: Looking to the 2LGBTQIA+ community to shape the future of sex robots. In *2022 17th ACM/IEEE International Conference on Human–Robot Interaction (HRI)* (pp. 746–749). IEEE. https://doi.org/10.1109/HRI53351.2022.9889580

Eagly, A. H., & Koenig, A. M. (2021). The vicious cycle linking stereotypes and social roles. *Current Directions in Psychological Science*, **30**, 343–350. https://doi.org/10.1177/09637214211013775

Eyssel, F., & Hegel, F. (2012). (S)he's got the look: Gender stereotyping of robots. *Journal of Applied Social Psychology*, **42**, 2213–2230. https://doi.org/10.1111/j.1559-1816.2012.00937.x

Eyssel, F., & Perugia, G. (2024) How people perceive social robots: The case of gender. In *The De Gruyter handbook of robots in society and culture*. De Gruyter.

Fiske, S. T. (1993). Controlling other people: The impact of power on stereotyping. *American Psychologist*, **48**, 621–628. https://doi.org/10.1037/0003-066X.48.6.621

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (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**, 878–902. https://doi.org/10.1037/0022-3514.82.6.878

Fogg, B. J. (1998, January). Persuasive computers: Perspectives and research directions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 225–232). ACM Press/Addison-Wesley Publishing Co. https://doi.org/https://dl.acm.org/doi/10.1145/274644.274677

Fraune, M. R., Šabanović, S., & Smith, E. R. (2020). Some are more equal than others: Ingroup robots gain some but not all benefits of team membership. *Interaction Studies*, **21**, 303–328. https://doi.org/10.1075/is.18043.fra

Galatolo, A., Melsión, G. I., Leite, I., & Winkle, K. (2023). The right (wo)man for the job? Exploring the role of gender when challenging gender stereotypes with a social robot. *International Journal of Social Robotics*, **15**, 1933–1947. https://doi.org/10.1007/s12369-022-00938-2

Glick, P., & Fiske, S. T. (1996). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology*, **70**, 491–512. https://doi.org/10.1037/0022-3514.70.3.491

Glick, P., & Fiske, S. T. (1999). The ambivalence toward men inventory: Differentiating hostile and benevolent beliefs about men. *Psychology of Women Quarterly*, **23**, 519–536. https://doi.org/10.1111/j.1471-6402.1999.tb00379.x

Golombok, S., & Fivush, R. (1994). *Gender development*. Cambridge University Press.

Goodman, K. L., & Mayhorn, C. B. (2023). It's not what you say but how you say it: Examining the influence of perceived voice assistant gender and pitch on trust and reliance. *Applied Ergonomics*, **106**, 103864. https://doi.org/10.1016/j.apergo.2022.103864

Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition*, **125**, 125–130. https://doi.org/10.1016/j.cognition.2012.06.007

Gross, N. (2023). What ChatGPT tells us about gender: A cautionary tale about performativity and gender biases in AI. *Social Sciences*, **12**, 435. https://doi.org/10.3390/socsci12080435

Guidi, S., Boor, L., van der Bij, L., Foppen, R., Rikmenspoel, O., & Perugia, G. (2022). Ambivalent stereotypes towards gendered robots: The (im)mutability of bias towards female and neutral robots. In F. Cavallo, J.-J. Cabibihan, L. Fiorini, A. Rossi, H. He, S. S. Ge, M. I. A. Ferreira, & R. Limosani (Eds.), *Social robotics* (pp. 615–626). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-24670-8_54

Hamidi, F., Scheuerman, M. K., & Branham, S. M. (2018, April). Gender recognition or gender reductionism? The social implications of embedded gender recognition systems. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1–13). Association for Computing Machinery. https://doi.org/10.1145/3173574.3173582

Hartvigsen, T., Gabriel, S., Palangi, H., Sap, M., Ray, D., & Kamar, E. (2022). Toxigen: A large-scale machine-generated dataset for adversarial and implicit hate speech detection, arXiv, arXiv:2203.09509. https://doi.org/10.18653/v1/2022.acl-long.234

Heitlinger, L., Stock-Homburg, R., & Wolf, F. D. (2022, March). You got the job! Understanding hiring decisions for robots as organizational members. In *2022 17th ACM/IEEE International Conference on Human–Robot Interaction (HRI)* (pp. 530–540). IEEE. https://doi.org/10.1109/HRI53351.2022.9889444

Hope, M., & Székely, É. (2025). Voices of 'cyborg awesomeness': Posthuman embodiment of nonbinary gender expression in AI speech technologies. In *26th Interspeech Conference 2025, Rotterdam, Netherlands, Kingdom of the, August 17-21, 2025* (pp. 689–693). International Speech Communication Association. https://doi.org/10.21437/Interspeech.2025-2229

Hsu, N., Badura, K. L., Newman, D. A., & Speach, M. E. P. (2021). Gender, "masculinity," and "femininity": A meta-analytic review of gender differences in agency and communion. *Psychological Bulletin*, **147**, 987–1011. https://doi.org/10.1037/bul0000343

Huang, R., Kim, M., & Lennon, S. (2022). Trust as a second-order construct: Investigating the relationship between consumers and virtual agents. *Telematics and Informatics*, **70**, 101811. https://doi.org/10.1016/j.tele.2022.101811

Hundt, A., Azeem, R., Mansouri, M., & Brandão, M. (2025). LLM-driven robots risk enacting discrimination, violence, and unlawful actions. *International Journal of Social Robotics*, **17**, 2663–2711. https://doi.org/10.1007/s12369-025-01301-x

Joel, D., Berman, Z., Tavor, I., Wexler, N., Gaber, O., Stein, Y., Shefi, N., Pool, J., Urchs, S., Margulies, D. S., Liem, F., Hanggi, J., Jancke, L., & Assaf, Y. (2015). Sex beyond the genitalia: The human brain mosaic. *Proceedings of the National Academy of Sciences*, **112**, 15468–15473. https://doi.org/10.1073/pnas.1509654112

Kachel, S., Steffens, M. C., & Niedlich, C. (2016). Traditional masculinity and femininity: Validation of a new scale assessing gender roles. *Frontiers in Psychology*, **7**, 956. https://doi.org/10.3389/fpsyg.2016.00956

Katz-Wise, S. L. (2020). *Gender fluidity: What it means and why support matters*. Harvard Health Blog. https://www.health.harvard.edu/blog/gender-fluidity-what-it-means-and-why-support-matters-2020120321544

Law, T., Chita-Tegmark, M., & Scheutz, M. (2021b). The interplay between emotional intelligence, trust, and gender in human–robot interaction: A vignette-based study. *International Journal of Social Robotics*, **13**, 297–309. https://doi.org/10.1007/s12369-020-00624-1

Law, T., Malle, B. F., & Scheutz, M. (2021a). A touching connection: How observing robotic touch can affect human trust in a robot. *International Journal of Social Robotics*, **13**, 2003–2019. https://doi.org/10.1007/s12369-020-00729-7

McDonnell, M., & Baxter, D. (2019). Chatbots and gender stereotyping. *Interacting with Computers*, **31**, 116–121. https://doi.org/10.1093/iwc/iwz007

Monahan, J. L. (1998). I don't know it but I like you: The influence of nonconscious affect on person perception. *Human Communication Research*, **24**, 480–500. https://doi.org/10.1111/j.1468-2958.1998.tb00428.x

Morgenroth, T., & Ryan, M. K. (2018). Gender trouble in social psychology: How can Butler's work inform experimental social psychologists' conceptualization of gender? *Frontiers in Psychology*, **9**, 1320. https://doi.org/10.3389/fpsyg.2018.01320

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, **56**, 81–103. https://doi.org/10.1111/0022-4537.00153

Neuteboom, S. Y., & de Graaf, M. (2021). Cobbler stick with your reads: People's perceptions of gendered robots performing gender stereotypical tasks, arXiv, arXiv:2104.06127. https://doi.org/10.1145/3434074.3447230

Nomura, T., & Suzuki, T. (2022). Relationships between humans' gender conception, expected gender appearances, and the roles of robots: A survey in Japan. *International Journal of Social Robotics*, **14**, 1311–1321. https://doi.org/10.1007/s12369-022-00873-2

Norman, D. (2007). *Emotional design: Why we love (or hate) everyday things*. Basic Books.

Norton, R. (2009, August 14). *Maiden names and little sports: The gay subculture in Georgian England*. http://rictornorton.co.uk/eighteen/maiden.htm

O'Connor, S., & Liu, H. (2023). Gender bias perpetuation and mitigation in AI technologies: Challenges and opportunities. *AI & Society*, **39**, 2045–2057. https://doi.org/10.1007/s00146-023-01675-4

Ostrowski, A. K., Walker, R., Das, M., Yang, M., Breazeal, C., Park, H. W., & Verma, A. (2022, August). Ethics, equity, & justice in human–robot interaction: A review and future directions. *In 2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)* (pp. 969–976). IEEE. https://doi.org/10.1109/RO-MAN53752.2022.9900805

Paetzel, M., Peters, C., Nyström, I., & Castellano, G. (2016, October). Effects of multimodal cues on children's perception of uncanniness in a social robot. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction* (pp. 297–301).

Association for Computing Machinery. https://doi.org/10.1145/2993148.2993157

Park, B., & Rothbart, M. (1982). Perception of out-group homogeneity and levels of social categorization: Memory for the subordinate attributes of in-group and out-group members. *Journal of Personality and Social Psychology*, **42**, 1051–1068. https://doi.org/10.1037/0022-3514.42.6.1051

Parlangeli, O., Palmitesta, P., Bracci, M., Marchigiani, E., & Guidi, S. (2023a). Gender role stereotypes at work in humanoid robots. *Behaviour & Information Technology*, **42**, 316–327. https://doi.org/10.1080/0144929X.2022.2150565

Parlangeli, O., Palmitesta, P., Masi, L., Tittarelli, M., & Guidi, S. (2023b). It's a long way to neutrality: An evaluation of gendered artificial faces. In *International Conference on Human–Computer Interaction* (pp. 366–378). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-35708-4_27

Perugia, G., Guidi, S., Bicchi, M., & Parlangeli, O. (2022). The shape of our bias: Perceived age and gender in the humanoid robots of the ABOT database. In *2022 17th ACM/IEEE International Conference on Human–Robot Interaction (HRI)* (pp. 110–119). IEEE. https://doi.org/10.1109/HRI53351.2022.9889366

Perugia, G., & Lisy, D. (2023). Robot's gendering trouble: A scoping review of gendering humanoid robots and its effects on HRI. *International Journal of Social Robotics*, **15**, 1725–1753. https://doi.org/10.1007/s12369-023-01061-6

Raghunath, N., Sanchez, C. A., & Fitter, N. T. (2022, December). Robot comedy (is) special: A surprising lack of bias for gendered robotic comedians. In *International Conference on Social Robotics* (pp. 663–673). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-24670-8_58

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people*. Vol. 10. Cambridge University Press.

Richardson, K. (2016). Sex robot matters: Slavery, the prostituted, and the rights of machines. *IEEE Technology and Society Magazine*, **35**, 46–53. https://doi.org/10.1109/MTS.2016.2554421

Robertson, J. (2010). Gendering humanoid robots: Robo-sexism in Japan. *Body & Society*, **16**, 1–36. https://doi.org/10.1177/1357034X10364767

Rollero, C., Peter, G., & Tartaglia, S. (2014). Psychometric properties of short versions of the ambivalent sexism inventory and ambivalence toward men inventory. *TPM. Testing, Psychometrics, Methodology in Applied Psychology*, **21**, 149–159. https://doi.org/10.4473/TPM21.2.3

Scheuerman, M. K., Paul, J. M., & Brubaker, J. R. (2019). How computers see gender: An evaluation of gender classification in commercial facial analysis services. *Proceedings of the ACM on Human–Computer Interaction*, **3**, 1–33. https://doi.org/10.1145/3359246

Schudson, Z. C., & Morgenroth, T. (2022). Non-binary gender/sex identities. *Current Opinion in Psychology*, **48**, 101499. https://doi.org/10.1016/j.copsyc.2022.101499

Scutella, M., Plewa, C., & Reaiche, C. (2022). Virtual agents in the public service: Examining citizens' value-in-use. *Public Management Review*, **26**, 73–88. https://doi.org/10.1080/14719037.2022.2044504

Seaborn, K., Barbareschi, G., & Chandra, S. (2023). Not only WEIRD but "uncanny"? A systematic review of diversity in human–robot interaction research. *International Journal of Social Robotics*, **15**, 1841–1870. https://doi.org/10.1007/s12369-023-00968-4

Seaborn, K., & Frank, A. (2022, April). What pronouns for pepper? A critical review of gender/ing in research. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (pp. 1–15). Association for Computing Machinery. https://doi.org/10.1145/3491102.3501996

Seaborn, K., Miyake, N. P., Pennefather, P., & Otake-Matsuura, M. (2021). Voice in human–agent interaction: A survey. *ACM Computing Surveys*, **54**, 1–43. https://doi.org/10.1145/3386867

Seaborn, K., & Pennefather, P. (2022a). Neither "hear" nor "their": Interrogating gender neutrality in robots. In *2022 17th ACM/IEEE International Conference on Human–Robot Interaction (HRI)* (pp. 1030–1034). IEEE. https://doi.org/10.1109/HRI53351.2022.9889350

Seaborn, K., & Pennefather, P. (2022b). Gender neutrality in robots: An open living review framework. In *2022 17th ACM/IEEE International Conference on Human–Robot Interaction (HRI)* (pp. 634–638). IEEE. https://doi.org/10.1109/HRI53351.2022.9889663

Søraa, R. A. (2017). Mechanical genders: How do humans gender robots? *Gender, Technology and Development*, **21**, 99–115. https://doi.org/10.1080/09718524.2017.1385320

Spiel, K. (2021, June). "Why are they all obsessed with gender?"—(Non)binary navigations through technological infrastructures. In *Designing Interactive Systems Conference 2021* (pp. 478–494). Association for Computing Machinery. https://doi.org/10.1145/3461778.3462033

Spinuzzi, C. (2005). The methodology of participatory design. *Technical Communication*, **52**, 163–174.

Steinhaeusser, S. C., Schaper, P., Bediako Akuffo, O., Friedrich, P., Ön, J., & Lugrin, B. (2021, March). Anthropomorphize me! Effects of robot gender on listeners' perception of the social robot nao in a storytelling use case. In *Companion of the 2021 ACM/IEEE International Conference on Human–Robot Interaction* (pp. 529–534). Association for Computing Machinery. https://doi.org/10.1145/3434074.3447228

Stolp-Smith, M., & Williams, T. (2024, March). More than binary: Transgender and non-binary perspectives on human robot interaction. In *Proceedings of the 2024 ACM/IEEE International Conference on Human–Robot Interaction* (pp. 697–705). Association for Computing Machinery. https://doi.org/10.1145/3610977.3634939

Sutton, S. J. (2020, July). Gender ambiguous, not genderless: Designing gender in voice user interfaces (VUIs) with sensitivity. In *Proceedings of the 2nd Conference on Conversational User Interfaces* (pp. 1–8). Association for Computing Machinery. https://doi.org/10.1145/3405755.3406123

Székely, É., & Hope, M. (2024). An inclusive approach to creating a palette of synthetic voices for gender diversity. In *Proceedings of Interspeech 2024* (pp. 3070–3074). International Speech Communication Association. https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1985230&dswid=-9677

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, **1**, 149–178. https://doi.org/10.1002/ejsp.2420010202

Tajfel, H., & Turner, J. C. (1979) An integrative theory of intergroup conflict. In W. G. Austin, & S. Worchel (eds), *The social psychology of intergroup relations*, pp. 33–37. Brooks/Cole.

Tay, B., Jung, Y., & Park, T. (2014). When stereotypes meet robots: The double-edge sword of robot gender and personality in human–robot interaction. *Computers in Human Behavior*, **38**, 75–84. https://doi.org/10.1016/j.chb.2014.05.014

Thorne, N., Yip, A. K. T., Bouman, W. P., Marshall, E., & Arcelus, J. (2019). The terminology of identities between, outside and beyond the gender binary – A systematic review. *International Journal of Transgenderism*, **20**, 138–154. https://doi.org/10.1080/15532739.2019.1640654

Tolmeijer, S., Zierau, N., Janson, A., Wahdatehagh, J. S., Leimeister, J. M. M., & Bernstein, A. (2021, May). Female by default? Exploring the effect of voice assistant gender and pitch on trait and trust attribution. In

Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1–7). Association for Computing Machinery. https://doi.org/10.1145/3411763.3451623

Torre, I., Lagerstedt, E., Dennler, N., Seaborn, K., Leite, I., & Székely, É. (2023, August). Can a gender-ambiguous voice reduce gender stereotypes in human–robot interactions? In *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)* (pp. 106–112). IEEE. https://doi.org/10.1109/RO-MAN57019.2023.10309500

Udry, J. R. (1994). The nature of gender. *Demography*, **31**, 561–573. https://doi.org/10.2307/2061790

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, **46**, 186–204. https://doi.org/10.1287/mnsc.46.2.186.11926

Verkuyten, M. (2021). Group identity and ingroup bias: The social identity approach. *Human Development*, **65**, 311–324. https://doi.org/10.1159/000519089

Weisman, H., & Janardhanan, N. S. (2020). The instantaneity of gendered voice assistant technology and manager perceptions of subordinate help. In *Academy of Management Proceedings* (p. 21149). Academy of Management. https://doi.org/10.5465/AMBPP.2020.21149abstract

Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin*, **25**, 1115–1125. https://doi.org/10.1177/01461672992512005

West, M., Kraut, R., & Ei Chew, H. (2019). *I'd blush if I could: Closing gender divides in digital skills through education*. UNESCO. https://doi.org/10.54675/RAPC9356

Winkle, K., Lagerstedt, E., Torre, I., & Offenwanger, A. (2023a). 15 years of (who) man robot interaction: Reviewing the h in human–robot interaction. *ACM Transactions on Human–Robot Interaction*, **12**, 1–28. https://doi.org/10.1145/3571718

Winkle, K., McMillan, D., Arnelid, M., Harrison, K., Balaam, M., Johnson, E., & Leite, I. (2023b). 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* (pp. 72–82). Association for Computing Machinery. https://doi.org/10.1145/3568162.3576973

Winkle, K., Melsión, G. I., McMillan, D., & Leite, I. (2021, March). Boosting robot credibility and challenging gender norms in responding to abusive behaviour: A case for feminist robots. In *Companion of the 2021 ACM/IEEE International Conference on Human–Robot Interaction* (pp. 29–37). Association for Computing Machinery. https://doi.org/10.1145/3434074.3446910

Winner, L. (2017) Do artifacts have politics? In *Computer ethics*, pp. 177–192. Routledge.

Ye, H., Jeong, H., Zhong, W., Bhatt, S., Izzetoglu, K., Ayaz, H., & Suri, R. (2020) The effect of anthropomorphization and gender of a robot on human–robot interactions. In H. Ayaz (ed), *Advances in neuroergonomics and cognitive engineering*, pp. 357–362. Springer.

Yu, C., Fu, C., Chen, R., & Tapus, A. (2022, March). First attempt of gender-free speech style transfer for genderless robot. In *2022 17th ACM/IEEE International Conference on Human–Robot Interaction (HRI)* (pp. 1110–1113). IEEE. https://doi.org/10.1109/HRI53351.2022.9889533

Zhang, Q., Yang, X. J., & Robert, L. P. (2025). Artificial intelligence voice gender, gender role congruity, and trust in automated vehicles. *Scientific Reports*, **15**, 16364. https://doi.org/10.1038/s41598-025-00884-9

Zhu, Y., Wen, R., & Williams, T. (2024, March). Robots for Social Justice (R4SJ): Toward a more equitable practice of human–robot interaction. In *Proceedings of the 2024 ACM/IEEE International Conference on Human–Robot Interaction* (pp. 850–859). Association for Computing Machinery. https://doi.org/10.1145/3610977.3634944

Interacting with Computers, 2026, 1–17
https://doi.org/10.1093/iwc/iwag030
Long article