*robotics*

MDPI

*Article*

# Nonbinary Voices for Digital Assistants—An Investigation of User Perceptions and Gender Stereotypes

**Sonja Theresa Längle, Stephan Schlögl *, Annina Ecker, Willemijn S. M. T. van Kooten and Teresa Spieß**

MCI—The Entrepreneurial School, Deptment of Management, Communication & IT, Universitätsstrasse 15, 6020 Innsbruck, Austria; willemijn.vankooten@mci.edu (W.S.M.T.v.K.); teresa.spiess@mci.edu (T.S.)
* Correspondence: stephan.schloegl@mci.edu; Tel.: +43-512-2070-3535

**Abstract:** Due to the wide adoption of digital voice assistants (DVAs), interactions with technology have also changed our perceptions, highlighting and reinforcing (mostly) negative gender stereotypes. Regarding the ongoing advancements in the field of human–machine interaction, a developed and improved understanding of and awareness of the reciprocity of gender and DVA technology use is thus crucial. Our work in this field expands prior research by including a nonbinary voice option as a means to eschew gender stereotypes. We used a between-subject quasi-experimental questionnaire study (female voice vs. male voice vs. nonbinary voice), in which $n = 318$ participants provided feedback on gender stereotypes connected to voice perceptions and personality traits. Our findings show that the overall gender perception of our nonbinary voice leaned towards male on the gender spectrum, whereas the female-gendered and male-gendered voices were clearly identified as such. Furthermore, we found that feminine attributes were clearly tied to our female-gendered voice, whereas the connection of masculine attributes to the male voice was less pronounced. Most notably, however, we did not find gender-stereotypical trait attributions with our nonbinary voice. Results also show that the likability of our female-gendered and nonbinary voices was lower than it was with our male-gendered voice, and that, particularly with the nonbinary voice, this likability was affected by people's personality traits. Thus, overall, our findings contribute (1) additional theoretical grounding for gender-studies in human–machine interaction, and (2) insights concerning peoples' perceptions of nonbinary voices, providing additional guidance for researchers, technology designers, and DVA providers.

**Keywords:** digital voice assistants; nonbinary voices; big five personality traits; gender stereotypes

check for updates

**Citation:** Längle, S.T.; Schlögl, S.; Ecker, A.; van Kooten, W.S.M.T.; Spieß, T. Nonbinary Voices for Digital Assistants—An Investigation of User Perceptions and Gender Stereotypes. *Robotics* **2024**, *13*, 111. https://doi.org/10.3390/robotics13080111

Academic Editors: Soyeon Caren Han and Naira Hovakimyan

Received: 5 June 2024
Revised: 16 July 2024
Accepted: 19 July 2024
Published: 23 July 2024

**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

## 1. Introduction

With 4.6 billion devices used, Digital Voice Assistants (DVAs) have become a broadly adopted consumer technology [1]. Approximately 36% of Americans own DVA-driven smart speakers, and half of them use these devices on a daily basis [2]. This amounts to over 1 billion voice searches conducted every month [3]. Apple's `Siri` and Google's `Assistant` hold the greatest market share, at roughly 36% each. Amazon's `Alexa` follows closely behind [4]. Other alternatives include Samsung's `Bixby`, Microsoft's `Cortana`, and a number of DVAs are offered by smaller companies. One characteristic all these DVAs have in common is their choice of gender representation. Not only are the names companies give to their DVAs predominantly female (i.e., Alexa, Siri, etc.), so are the voices they equip those artificial agents with, which convey a stereotypical gender role. While this may seem trivial, it goes to the heart of how the spread of digital technology can amplify and potentially even extend both gender stereotypes and social prejudice. To this end, a report published by UNESCO clearly points out that there is a danger of DVAs becoming modern instruments for gender bias manifestation [5]; in particular, since humans tend to attribute human characteristics or behavior to technology in general [6], and even more so to technology that is considered anthropomorphized, such as DVAs [7]. Thus, although `Siri` and `Alexa` might

reply that they are genderless like cacti and certain types of fish when asked about their gender identity, technology designers still create these digital entities with a female persona in mind, equipping them with specific female gender cues and consequently triggering gender-biased user perceptions. Because, at present, users predominantly interact with female-gendered voice assistants, it is worth reflecting upon how these DVA interactions might manifest outdated ideas about women's roles in society, and what may be done to avert such stereotyping. Furthermore, although today's DVA settings usually allow for a binary choice of gender, i.e., male or female, the fluid concept of gender suggests that it is a spectrum rather than a misconceived and often simplified binary option [8], calling for more variance in addressing the complexities of gender.

DVA characteristics are highly dependent on human designers, and thus there is a certain risk that prevailing human biases are carried over to AI-driven systems. Yet, this risk, if adequately addressed, may also present an opportunity, in that diversity and inclusiveness could become an inherent part of DVA technology design and development. For DVA design, this would mean the promotion of more gender-neutral DVA names and, additionally, more investment into building respective gender-neutral DVA voices. That is, next to (or even instead of) male and female voices, users should be able to choose from a selection of nonbinary voices, which would give them the chance to interpret the voice—either as female, male, or neither, making it about choice rather than conditioning.

So far, however, our knowledge of users' perception regarding a third, nonbinary gender voice option in DVAs is still rather limited (note: the voice `Q` promoted by GenderlessVoice is not currently implemented in any of the commercially available DVAs; available online: https://genderlessvoice.com/ [accessed on 4 June 2024]). The work presented in this article thus aims to take a first step towards a better understanding of this field by reporting on a study investigating users' DVA voice perceptions and respective gender stereotype mapping. In particular, we investigated how individual factors—such as gender, education, profession, personality, and affinity for technology—relate to preferences for and stereotypes associated with nonbinary versus binary (male or female) voices.

The respective investigation was guided by the following research question:

> *How do people perceive nonbinary digital voices, and to what extent do they elicit gender stereotypes?*

The description of our investigation starts with a discussion of relevant theoretical concepts and related work in Section 2. Next, Section 3 describes our methodological approach, the materials and the instruments we used to tackle the above stated research question. Then, we report on the generated results in Section 4, and discuss their impact in Section 5. Finally, Section 6 concludes with limitations and potential areas for future investigation.

## 2. Theoretical Concepts and Related Work

Building machines that understand and respond to spoken language has always been a key goal of AI [9]. In this, the aim was not only to offer a more natural way of information exchange, but also to allow for the creation of deeper human–technology relationships [10]. As a result, we see how today's DVAs and their underlying spoken dialog system technology increasingly take over and further advance the role of hitherto used graphical user interfaces [11].

To this end, McTear [12] defines three main types of spoken dialogue systems used to power DVAs, i.e., finite state-based systems, frame-based systems, and agent-based systems. Finite state-based systems have strict rules that let interlocutors proceed step-by-step through a pre-defined dialog structure. They are rather inflexible, and consequently best suited for clearly defined, linear interaction scenarios. Most modern voice assistants, such as Apple's `Siri`, Amazon's `Alexa`, or Google's `Assistant`, therefore tend to employ the more flexible frame-based approach, in which tasks are completed by randomly filling in slots of task templates [13]. Finally, agent-based systems are considered the most flexible ones. They act as autonomous entities aimed at reaching their goals by sensing input and triggering adequate output/actions, where the level of adequacy is measured in the

closeness an output/action brings them to a desired goal [14]. In other words, an agent-based system runs on a continuous cycle of perceiving, processing (i.e., selecting the next action) and acting. Furthermore, these DVAs are also able to interact with other agents via an agent-communication language [15,16], which makes them versatile in that they can be integrated with a multitude of tasks and services.

### *2.1. Voice-Based Human–Machine Interaction*

Voice is considered a key modality in successful human–human interaction, thus making it also a suitable candidate for more intuitive human–machine interaction [17,18]. Furthermore, studies have shown that in human–machine interaction users often apply the same social rules to machines as they do to other people [11]. In other words, people have a tendency to see human-like qualities in non-human objects if those objects exhibit a certain level of anthropomorphism [19]. This also affects overall machine perception, and furthermore increases technology acceptance [20]. Consequently, when human–machine conversations feel successful and cooperative, users may change their pronunciation, word choice, and other speech patterns so as to match human–human conversation patterns. Modern DVAs have already adapted to this behavior, and thus allow users to speak naturally as they would with other people instead of asking them to memorize DVA-specific trigger words or task commands [21].

From an interaction point of view, scholars have also pointed to a certain similarity-attraction effect in human–human and as well as human–machine interaction [10,22]. That is, people tend to be attracted to individuals (or artificial entities for that matter) who are like themselves. The attraction is thereby not limited to physical attractiveness, but rather the desire to be around such a person [23], which may be triggered by, e.g., vocal beauty [24]. This also applies to technology exhibiting anthropomorphic characteristics, such as DVAs. Thus, although personalized user interfaces are not new [25], the focus on individualized experiences is particularly present in modern DVAs [15]. On the flip side, however, this also opens the door for unconscious manifestations of already existing societal stereotypes—particularly those related to gender.

### *2.2. The Gender Binary*

The framework of the gender binary refers to the concept of individuals being either men or women. Both internal (biological, cognitive) and external (social interaction, culture) mechanisms contribute to this view of gender. Cole [26], however, argues that this binary view fails to address the complexity of gender. Furthermore, it excludes gender-nonconforming individuals by simply denying their existence. People's gender and their gendered traits should thus rather be considered the "*intended or unintended product[s] of a social practice*" ([27], p. 97). To this end, gender performance theory proposes a clear distinction between sex and gender, pointing out that masculinity and femininity are not identical to men and women [8]. Gender is rather socially constructed, and defined by cultural norms. Evidence of this manifests in expectations among different cultures, where time is considered a factor that influences the meaning of gender and its changing nature [28]. Consequently, the "*limits [of gender] are always set within the terms of a hegemonic cultural discourse predicated on binary structures that appear as the language of universal rationality*" ([8] p. 12). So, while one may not completely neglect the gender binary, it should at least be acknowledged that the concept of gender is dynamic and responsive [29]. With respect to anthropomorphic technology, it was thus suggested to focus more on androgynous humanoid characteristics, and move away from the more traditional stereotyped gender representations [28]. An example of this may be found in the humanoid robot `Pepper`, where the developers explicitly omitted the use of gender-typical characteristics. Although only 15 of 50 study participants correctly classified `Pepper` as gender-neutral [30] (note: the majority, i.e., 64% perceived the robot as male, and 6% of study participants classified it as a female robot), it may still be considered a significant step towards overcoming the gender binary inherent to anthropomorphic agent technology.

### 2.3. Gender and Digital Voice Assistants

While DVAs lack an inherent biological gender, they often acquire a social-mechanical gender through human interaction and design [31]. This gender assignment is influenced by societal perceptions and stereotypes about male and female roles, leading to criticism regarding the reinforcement of harmful biases [32]. These stereotypes, based on societal expectations and assumptions about gender roles, influence interactions between humans and seemingly human-like entities [33]. For instance, people tend to associate male DVAs with authority and female DVAs with kindness, mirroring common gender stereotypes [34]. The context in which a DVA is used also impacts the assigned gender stereotypes. Here, research has shown that perceived task suitability influences the gender stereotypes associated with an artificial entity. So is it that female-presenting chatbots, for example, receive more comments on physical appearance than male-presenting ones [35]. This integration of gender characteristics into DVAs perpetuates traditional gender norms and biases, shaping our understanding of gender roles [28,36]. It stems from existing societal biases [37], and thus reinforces existing discrimination against women. For example, DVAs predominantly performing service-oriented tasks reinforces the stereotype of women in subservient roles [38]. Furthermore, the design of DVAs is inevitably influenced by the implicit values and biases of their human creators [39,40], which not only raises concerns about the devaluation of women's work, but also the reinforcement of subordinate roles for women in technology [37].

Designing voice interface interactions requires consideration of users, devices, and contextual factors [22,33]. To this end, a user's gender significantly influences technology adoption and perception of DVAs. Here, studies have shown gender differences in new technology adoption, with males being more influenced by personal attitudes and females by perceived norms and control [41]. That is, user responses to DVAs vary depending on their own gender identity and the perceived gender of the DVA [42]. Similarly, research indicates gender differences in perceiving human-like characteristics in robots, with males often attributing more human-like qualities and displaying a more positive attitude towards robots compared to females [43–45]. Yet, this can lead to varying behavioral responses, such as donation behavior influenced by the perceived gender of the artificial entity [46].

These findings highlight the need for a deeper understanding of social cues impacting DVA acceptance and addressing the ethical implications of anthropomorphized technology [47]. Only recently, developments in DVA design indicate a shift towards more neutral DVA interactions, which potentially move away from reinforcing stereotypes and towards creating more inclusive representations of gender in said technology [48]. This type of technology design, which promotes social equality and challenges stereotypes, is crucial to avoid the reinforcement of harmful biases [49,50]. It can only be achieved by creating artificial identities that transcend social stereotypes and spark conversations about the representation of gender in technology [51].

### 2.4. Promoting a Nonbinary Option for Digital Voice Assistants

While individuals may self-identify as binary (male or female) or nonbinary (neither male nor female) [35], there are external characteristics that (intentionally or unintentionally) convey a certain gender identity to others. With DVAs, this can happen directly through the content of a conversation in which a DVA may explicitly state to be male, female, or nonbinary, or indirectly through the used voice, which implicitly carries gender identity information. One such voice characteristic, which differentiates female and male DVA identities, is the so-called Speaking Fundamental Frequency (SFF) $F_0$ and its parameters, i.e., range and variability. A gender-neutral SFF range, for example, would be found between $F_0$ = 145–175 Hz [52,53], whereas an adult female SFF would average at approximately $F_0$ = 220 Hz and an adult male SFF at approximately $F_0$ = 120 Hz [54,55]. But it is not the SFF alone that controls our gender perception. Additionally, it has been shown that inflection, voice quality, articulation, speaking intensity, speech rate, and prosody matter [52]. In all of this, nonbinary voices should neither pattern like women nor like men. Rather, they

should use a combination of male and female characteristics [53]. As a nonbinary voice lies between the parameters of female and male voices, it can potentially mitigate gender stereotyping. Thus, modifying the tonality from a typically female-gendered voice to a nonbinary voice may positively affect gender bias [56]. Aiming to challenge the prevalent binary representations in synthesized voices, the [multi'vocal] project (available online: https://thenewnew.space/projects/multivocal/ [accessed on 4 June 2024]) engaged a diverse set of people across different genders, ages, and accents to create a nonbinary synthesized voice [57]. It may be considered an example of how gender-neutral design efforts can help guard against unconscious stereotypes perpetuating gender inequalities and might motivate users to rethink and eventually reconfigure socially constructed gender norms [49]. Unfortunately, however, we currently know little about people's perception of such nonbinary voices in DVAs and how this impacts on their gender stereotyping.

## 3. Methodology

Aiming to address the above outlined knowledge gap concerning nonbinary DVAs, we designed a quasi-experimental questionnaire study that investigates people's perceptions connected to differently gendered DVA voices. Within this study, we considered the interplay of gender and technology using a two-tier quantitative research approach. Firstly, we investigated the general perception of binary and nonbinary DVA voices and their elicitation of gender stereotypes. Secondly, we analyzed the likability of these DVA voices, which we considered to be closely connected to users' voice preferences. We used German-speaking DVA users as our sample frame.

### *3.1. Research Model*

Our goal was to understand how people's personal characteristics affect their perception and, consequently, the likability of nonbinary DVA voices. Personal characteristics we considered relevant included people's gender identity, personality, egalitarianism, age, and education. With respect to the DVA voice, we expected its adopted gender to elicit corresponding gender stereotypes. That is, a female, male, or nonbinary voice was hypothesized to be connected to different stereotypical traits. The proposed research model depicted in Figure 1 summarizes these connections and outlines a number of hypotheses, which will be described next.

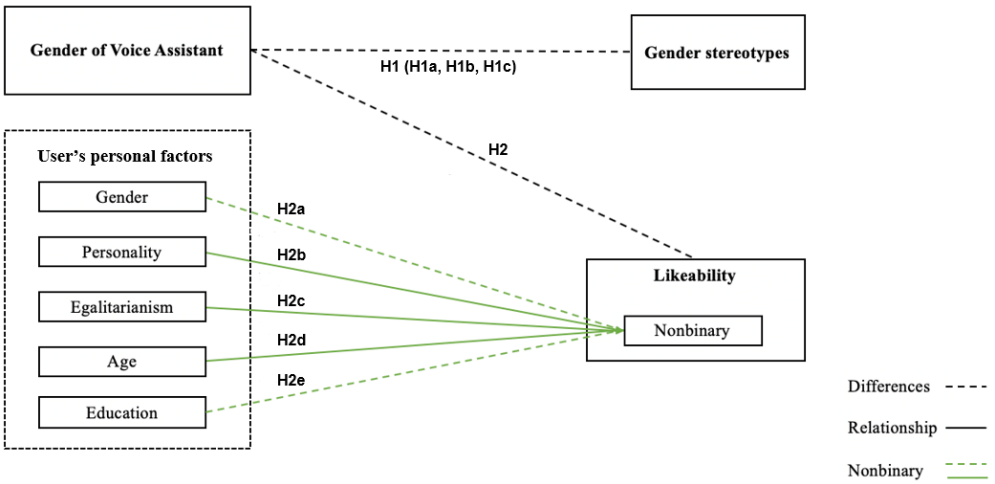



**Figure 1.** Proposed research model.

### *3.2. Hypotheses*

Nass et al. [32] found that people considered low-pitched voices male and high-pitched voices female. Consequently, we may assume that people assign different gender-stereotypical trait attributions to male, female, and nonbinary voices. We therefore hypothesize that:

**H1:** *In terms of gender-stereotypical trait attribution, there is a significant difference between female, male, and nonbinary voice assistants.*

In prior work, scholars have shown that the perceived gender elicits gender stereotypes. Female-gendered agents have been ascribed attributes such as friendly and polite, whereas male-gendered agents are perceived as authoritarian and dominant [36]. Consequently, we may argue that:

**H1a:** *People ascribe stereotypical feminine attributes to female-gendered voice assistants.*

**H1b:** *People ascribe stereotypical masculine attributes to male-gendered voice assistants.*

Furthermore, as nonbinary voices are not ascribed to the gender binary, they should not correspond with male or female gender stereotypes, supporting the assumption that:

**H1c:** *People do not assign stereotypical gendered attributes to nonbinary voice assistants.*

As outlined in Section 2, current commercially available DVAs are predominantly female-gendered. Companies consistently argue that this corresponds to people's preferences. Consequently, we may assume that:

**H2:** *There is a significant difference in the likability between female, male, and nonbinary voice assistants.*

Previous work has furthermore shown that personal characteristics affect the likability of DVAs. Respective studies draw on the social identity theory as a foundation to explain differences among user's gender identities [58], where the majority of the work indicates differences in people's DVA preferences depending on whether they identify as male or female [22,59,60]. Thus, we may hypothesize:

**H2a:** *There is a significant difference in the likability of nonbinary voice assistants between men and women.*

As DVA's personalization advances, the interest in people's personality and how it affects the likability of DVAs grows [15,61]. Using a short 10-item personality test, we thus set out to understand the relationship between different personality traits and the likability of nonbinary voice assistants, proposing that:

**H2b:** *There is a relationship between the likability of nonbinary voice assistants and a user's personality traits.*

Furthermore, following Moskowitz and Li [62], who identified that egalitarian goals trigger stereotype inhibition, we may assume that:

**H2c:** *There is a relationship between the likability of nonbinary voice assistants and a user's exhibited egalitarianism.*

Also, it is suggested that age is a factor influencing users' stereotype perceptions in DVAs [63], for which we hypothesize that:

**H2d:** *There is a relationship between the likability of nonbinary voice assistants and a user's age.*

Finally, education was shown to affect users' impressions towards differently gendered agents [64], leading to the assumption that:

**H2e:** *There is a significant difference in the likability of nonbinary voice assistants between different educational backgrounds.*

### *3.3. Materials*

We used an online questionnaire in which we let participants listen to two different DVA prompts uttered by one of three DVA voices (i.e., male, female, or nonbinary), and subsequently asked them questions related to the heard voice. The German prompts were situated in an e-commerce context, where the first one illustrated a system's response to a user aiming to purchase a text processing software and the second one to a user's search for a pen (cf. Table 1). The text for both response prompts was taken from an actual interactions with one of the commercially available DVAs.

**Table 1.** E-commerce-based DVA prompts uttered by either a female, male, or nonbinary voice as part of the online questionnaire.

| Product | Envisioned User Prompt | DVA Response Uttered by a Female, Male or Nonbinary Voice |
|---|---|---|
| Software | Ich will ein Textverarbeitungsprogramm kaufen. (en: I want to buy a text processing software.) | Ein Topergebnis ist Microsoft Office 365 Home multilingual Jahresabonnement sechs Nutzer Box. Der Preis beträgt neunundneunzig Euro und neunundneunzig Cent inklusive deutscher Mehrwertssteuer mit Lieferung bis zwanzigsten Mai. (en: A top result is Microsoft Office 365 Home multilingual annual subscription six user box. The price is ninety-nine euros and ninety-nine cents including German VAT with delivery by the twentieth of May.) |
| Pen | Ich will einen Kugelschreiber kaufen. (en: I want to buy a pen.) | Ein Topergebnis ist Faber Castell Kugelschreiber Poly Ball XB schwarz Schreibfarbe blau. Der Preis beträgt vier Euro und neunundachzig Cent inklusive deutscher Mehrwertssteuer. Wird voraussichtlich am zwanzigsten Mai geliefert. (en: A top result is Faber Castell ballpoint pen Poly Ball XB black writing color blue. The price is four euros and eighty-nine cents including German VAT. Expected to be delivered on the twentieth of May.) |

The version that included the nonbinary voice was distributed twice as often as the other two genders, while the female and male version served as controls. For the male and female voices, we opted for currently available state-of-the-art German synthesized options generated by the Google Cloud Platform. For the nonbinary option, we asked the Acapela group (available online: https://www.acapela-group.com/ [accessed on 4 June 2024]) to generate a dedicated German nonbinary voice for us. Their approach was to use a female voice as the starting point and then change the $F_0$ (fundamental frequency) and the ratios of the vowel formants $F_1$ (vowel height; e.g., i = high vs. a = low) and $F_2$ (vowel backness; e.g., i = front vs. u = back), so as to create a voice that has characteristics that lie between those of a male and a female voice. In a pre-study, we then asked five of our students to rate a curated selection of such nonbinary voices according to neutrality, on a gender spectrum from male to female, and naturalness, from computer-like to human-like (note: we were not interested in whether they liked the voice, but rather whether the voice would sound natural and would score on the gender spectrum in between male and female). We then opted for the one voice that scored best along these dimensions (note: the male and female voices did not undergo such pre-testing, since we assumed that current commercially available voices have already been tested for their gender representation).

### 3.4. Measures

In order to implement the research model illustrated in Figure 1 and investigate on its respective hypotheses, we used previously validated variable constructs derived from the literature. As we wanted to focus on German-speaking DVA users, we either used the validated German version of these measures or translated items via translation–back translation. The following variables and variable constructs were used (note: the entire questionnaire including a description of all data items and their parameter values is publicly available via the Zenodo open data directory at https://doi.org/10.5281/zenodo.11468262 [accessed on 7 July 2024]).

#### 3.4.1. Perceived Voice Gender

We measured the perceived voice gender in the same way as in the pre-study by asking participants to rate the perceived voice on a nine-point gender spectrum running from male (=1) to female (=9), with the middle point (=5) representing the nonbinary neutral position.

#### 3.4.2. Likability of Voices

Likability defines how much participants like the female, male, or nonbinary voices. We aimed for a multi-item measurement instrument that evaluates the likability of a perceived stimulus. Evaluating measures that have been used by other related work, we chose a scale proposed by Monahan [65]. It uses five items, which are each scored on a five-point semantic differential, and have previously been used to measure the likability of anthropomorphic voice user interfaces [20] as well as robots [66].

#### 3.4.3. Bem Sex Role Inventory

We assessed gender-stereotypical perceptions using indicatives for stereotypical masculine and feminine traits. In this, we referred to the classic research by Bem [67], who categorized stereotypical male and female attributes. Thus, a seven-point Likert scale was used to capture the intensity of each of the 12 attributes, and how they apply to the respective gendered DVA. Female-gendered agents are therein stereotypically ascribed traits related to interpersonal warmth, whereas male-gendered agents are connected to agentic features [36].

#### 3.4.4. Big Five Inventory

Respondents' personality traits are needed to investigate hypothesis H2b. We used the shortened 10-item Big Five Inventory (BFI-10) proposed by Rammstedt and John [68], as it has shown to be almost as reliable and valid as the much longer 44-item version of the BFI.

#### 3.4.5. Gender Role Stereotype Scale

The Gender Role Stereotype Scale (GRSS) was used to investigate potential gender role stereotypes based on eight items. Respondents were asked to indicate the extent to which different tasks should be accomplished by men, women or equally shared [69]. This measure relates to the egalitarianism hypothesis (H2c).

#### 3.4.6. Affinity for Technology Interaction

Finally, we used the nine-item Affinity for Technology Interaction (ATI) [70] scale to measure the extent to which a respondent engages in intensive technology interaction activities. Although not attached to any of our hypotheses, this helped us with the description of our sample and the consequent identification of those respondents who express a greater liking of technology and, thus, probably exhibit a better understanding of DVA technology.

### 3.5. Procedure

First, we asked respondents to listen to the two voice prompts provided by the questionnaire. Next, so as to control for potential technical errors, they had to confirm that they were able to listen to the prompts, and that they understood the therein presented content. Then, they had to evaluate the perceived DVA voice, based on the measures described above. Finally, we asked them to include standardized sociodemographic data, which we then used to characterize our sample. The survey was distributed online for two weeks within our personal and university networks, different social media groups, and online platforms, which ensured a rather diverse sample of respondents.

### 3.6. Statistical Analyses

Next to descriptive analyses employing measures of central tendency, i.e., mean (M), minimum values (min), maximum values (max), standard deviation (SD) and group frequencies, we used Cronbach's $\alpha$ to evaluate the internal consistency of all latent variable constructs, and Levene's test for equality of variance, as well as the Shapiro–Wilk test for normality to check for adequate value distributions. We then employed independent samples *t*-tests and analyses of variance (ANOVA), as well as post hoc analyses (e.g., Tukey test) to investigate differences (H1, H1a–c, H2, H2a, H2e) between two (*t*-test) or more groups (ANOVA). To further measure the effect size of possible differences between groups, we used Cohen's *d* (independent samples *t*-tests) and *f* (ANOVA) statistic. As for potential relationships between constructs and/or demographics (H2b–d), we used correlation and regression analyses. Finally, to provide an estimation of the affinity for technology present in our sample, we used the mean over the nine-item ATI-scale (note: the ATI items 3, 6 and 8 were reverse-coded to adjust item polarity).

## 4. Results

We received a total of $n = 318$ ($n = 168$ female and $n = 150$ male) valid responses to the questionnaire, distributed over the three different DVA voices (i.e., male voice = 76, female voice = 81 and nonbinary voice = 161). Respondents' mean age when completing the questionnaire was 26.75 years (min = 19, max = 66, SD = 7.25) (cf. Figure 2), and they resided mostly in Austria (73.3%) and Germany (22.6%). Other countries include Switzerland, Italy, France, and Liechtenstein (note: while some of the participants lived in non-German speaking countries, all of them were native German speakers). Most respondents hold a high school diploma (35.22%), bachelor's degree (41.19%) or master's degree (13.52%). Other educational backgrounds include a completed apprenticeship, a vocational baccalaureate diploma, a secondary education certificate, or other (e.g., a doctoral degree).

Approximately a third of the respondents (31.8%) claimed that they had never used or tried to use a DVA preceding their participation in our study, whereas 82 respondents (25.8%) stated they had used it once. The rest of the respondents indicated to use DVAs once or several times per month, once or several times per week, or daily (cf. Figure 3). Concerning respondents' affinity for technology, the results of the nine-item ATI scale point to a homogeneous sample exhibiting an average overall technology affinity ($M = 3.67$, $SD = 0.70$) (note: as stated before, we did not attach a hypothesis to the found level of technology affinity, but rather used it as a means to describe the collected sample).

As for respondents' occupations, the majority of them indicated being students (55.97%) or employees (33.96%). Thirteen people stated being self-employed, four people were in education or unemployed/looking for a job, and two were government officials (i.e., civil servants). The remaining 2.83% of respondents were either on parental leave or combining their studies with some sort of part-time work when they completed the questionnaire.

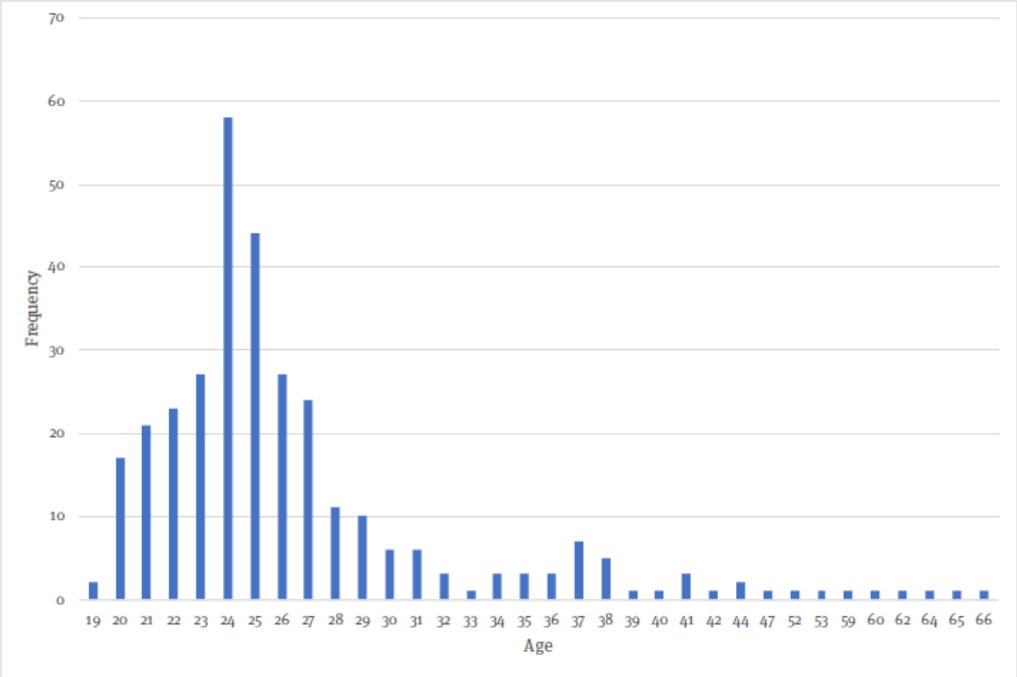



**Figure 2.** Respondents' age distribution.

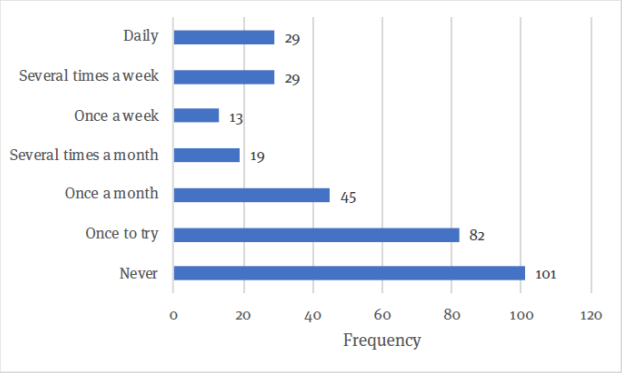



**Figure 3.** Respondents' stated frequency of using digital voice assistants.

### *4.1. Perceived Gender*

As outlined earlier, voice gender perception was measured on a nine-point gender spectrum running from male = 1 to female = 9, with the middle point = 5 representing the nonbinary neutral position. We analyzed responses regarding perceived gender using a one-way analysis of variance (ANOVA), to ensure that participants had correctly recognized the gendered voices as such. The results indicate significant differences in perceived gender between the male ($M = 1.76$, $SD = 0.99$), female ($M = 8.58$, $SD = 0.99$), and nonbinary voices ($M = 3.14$, $SD = 2.09$), $F(2, 315) = 404.369$, $p < 0.005$. The post hoc Tukey test shows a significant ($p = 0.000$) difference for all three genders, indicating that participants had a distinctively different perception of all three voices.

Comparing the DVA's intended gender representation with the gender perceived by the respondents shows that they identified the binary options with high accuracy (cf. Table 2). For the nonbinary voice, the majority of respondents (65.2%) perceived the voice as male, whereas only 24.8% of them correctly identified the nonbinary voice as neutral. A standard deviation of $SD = 2.09$ on the 1–9 gender spectrum, however, indicates a rather broad range of data points. In comparison, the standard deviations of the male ($SD = 0.992$) and female voice assistant ($SD = 0.986$) are smaller, indicating a

greater ambiguity in perceiving the nonbinary voice compared to perceiving the female and male voices.

**Table 2.** Perceived gender vs. actual gender of the DVA voices.

| | | Perceived Gender | | | |
|---|---|---|---|---|---|
| | | **Female** | **Male** | **Nonbinary** | **Correctly Identified** |
| *Actual Gender* | Female | 79 | 1 | 1 | 97.5% |
| | Male | 0 | 71 | 5 | 93.4% |
| | Nonbinary | 16 | 105 | 40 | 24.8% |

### *4.2. Elicitation of Gender Stereotypes (H1)*

To understand if people ascribe stereotypical gender traits to gendered voices, we ran a one-way ANOVA comparing female, male, and nonbinary voices with the respective feminine and masculine traits assigned by respondents. As outlined in Section 3.4.3, all traits were measured on a seven-point Likert scale running from 1 = *not true at all* to 7 = *totally true*, with 4 representing a neutral point. Internal consistency (Cronbach's $\alpha$) for masculine traits and feminine traits was calculated for both binary genders and confirm the reliability of the construct for feminine (female voice = 0.843, male voice = 0.787) as well as masculine traits (female voice = 0.763, male voice = 0.646) [71]. Due to the ambiguity of its perception, we did not expect the nonbinary voice's internal consistency to be a reliable indicator of the validity of the construct. Still, results indicate that at least two voices, i.e., the male and the female voice, score significantly differently for both feminine ($F(2, 315) = 33.824$, $p = 0.000$) and masculine ($F(2, 315) = 66.797$, $p = 0.000$) trait attributions (cf. Table 3).

**Table 3.** Differences in gender-stereotypical trait attribution based on perceived DVA voices.

| | | **Sum of Squares** | **df** | **Mean Square** | **F** | **Sig.** |
|---|---|---|---|---|---|---|
| **Feminine traits** | Between Groups | 65.5780 | 2 | 32.7890 | 33.8244 | 0.000 |
| | Within Groups | 305.3580 | 315 | 0.9694 | | |
| | Total | 370.9360587 | 317 | | | |
| **Masculine traits** | Between Groups | 83.1186 | 2 | 41.5593 | 66.7972 | 0.000 |
| | Within Groups | 195.9841 | 315 | 0.6222 | | |
| | Total | 279.1027254 | 317 | | | |

While masculine traits show a medium effect ($f = 0.5779$), feminine traits show a small effect ($f = 0.4552$) [72]. The post hoc Tukey test shows a significant difference ($p < 0.05$) between all genders for both gendered traits with one exception: there was no significant difference between the ascription of feminine traits to the male and nonbinary voice. Consequently, our data only partly supports H1.

Furthermore, we ran an independent samples *t*-test to compare the differences between the masculine and feminine trait attribution for each gender (female, male, and nonbinary). The results indicate that the mean feminine trait attribution is highest for the female voice. Similarly, the male voice scores the highest for masculine trait attribution. With the nonbinary voice, feminine and masculine traits are attributed almost equally. The results are illustrated in Figure 4 and further discussed below.

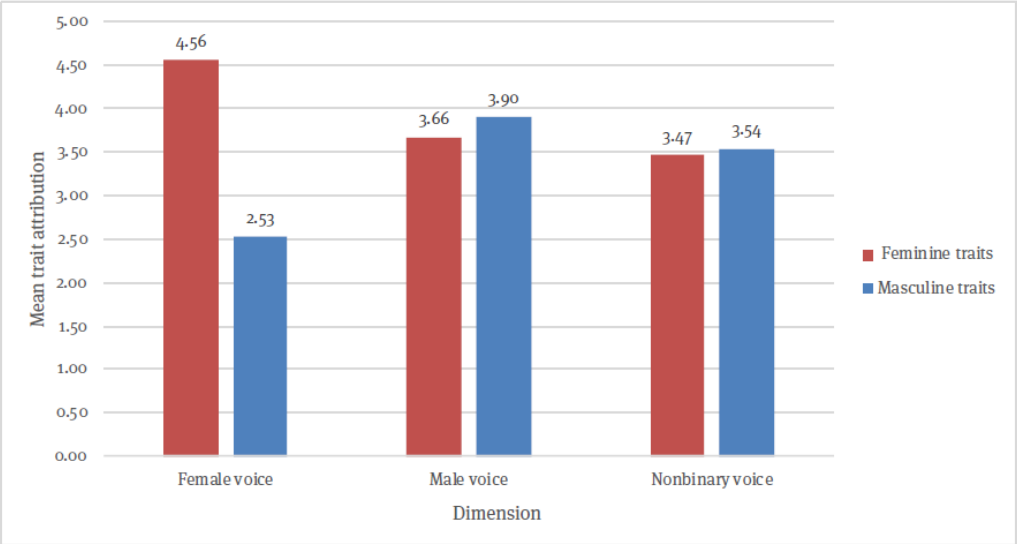



**Figure 4.** Gender-typical trait attribution based on perceived DVA voices.

#### 4.2.1. Female Voice (H1a)

The female voice scores a mean of 4.56 in feminine traits and a mean of 2.53 in masculine traits. An independent samples *t*-test for equality of means shows a significant difference ($t(160) = 13.706$, $p = 0.000$) in the feminine and masculine traits when the DVA's gender is female (cf. Table 4). Thus, a female DVA elicits statistically more feminine traits than masculine traits, supporting the assumption that gender stereotyping applies to female-gendered DVAs. Consequently, the collected data confirms our hypothesis H1a.

**Table 4.** Independent samples *t*-test: female voice.

| | | Levene's Test for Equality of Variances | | *t*-Test for Equality of Means | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | F | Sig. | t | df | Sig. (2-Tailed) | Mean Difference | Std. Error Difference | 95% Confidence Interval of the Difference | |
| | | | | | | | | | Lower | Upper |
| Gender-stereotypical traits | Equal variances assumed | 6.149 | 0.014 | 13.706 | 160 | 0.000 | 2.029 | 0.148 | 1.736 | 2.321 |
| | Equal variances not assumed | | | 13.706 | 152.244 | 0.000 | 2.029 | 0.148 | 1.736 | 2.321 |

#### 4.2.2. Male Voice (H1b)

The male voice shows a higher mean in masculine traits ($M = 3.90$) than in feminine traits ($M = 3.66$). Yet, even though the male voice scores, on average, are higher in masculine traits, the independent samples *t*-test does not show a statistically significant differentiation ($p = 0.085$) (cf. Table 5). As such, the results do not support our hypothesis H1b, stating that the male voice elicits masculine trait attributes.

**Table 5.** Independent samples *t*-test: male voice.

| | | Levene's Test for Equality of Variances | | *t*-Test for Equality of Means | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | F | Sig. | t | df | Sig. (2-Tailed) | Mean Difference | Std. Error Difference | 95% Confidence Interval of the Difference | |
| | | | | | | | | | Lower | Upper |
| Gender-stereotypical traits | Equal variances assumed | 2.331 | 0.129 | −1.733 | 150 | 0.085 | −0.243 | 0.140 | −0.520 | 0.034 |
| | Equal variances not assumed | | | −1.733 | 139.936 | 0.085 | −0.243 | 0.140 | −0.521 | 0.034 |

#### 4.2.3. Nonbinary Voice (H1c)

Finally, both the feminine and masculine mean trait attribution values are similar for the nonbinary DVA voice. Indeed, the *t*-test does not show a significant difference ($p = 0.492$) between feminine ($M = 3.47$) and masculine traits ($M = 3.54$) (cf. Table 6). We may therefore argue that people did not assign the nonbinary DVA voice to a specific gender and, consequently, that the nonbinary DVA voice does not elicit gender-stereotypical trait attributions, supporting our hypothesis H1c.

**Table 6.** Independent samples *t*-test: nonbinary voice.

| | | Levene's Test for Equality of Variances | | *t*-Test for Equality of Means | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | F | Sig. | t | df | Sig. (2-Tailed) | Mean Difference | Std. Error Difference | 95% Confidence Interval of the Difference | |
| | | | | | | | | | Lower | Upper |
| Gender-stereotypical traits | Equal variances assumed | 4.920 | 0.027 | −0.687 | 320 | 0.492 | −0.067 | 0.098 | −0.260 | 0.125 |
| | Equal variances not assumed | | | −0.687 | 308.748 | 0.492 | −0.067 | 0.098 | −0.260 | 0.125 |

### *4.3. Factors Influencing Likability (H2)*

The data show that the likability for DVAs is internally consistent (Cronbach's $\alpha = 0.898$), indicating a reliable construct [71]. Furthermore, the Shapiro–Wilk test ($p > 0.05$) shows that the likability for the nonbinary DVA is normally distributed. The one-way ANOVA points to a statistically significant difference in the likability of DVAs connected to their gender, confirming our hypothesis H2: $F(2, 315) = 12.736$, $p = 0.000$. On closer look, the post hoc Tukey test revealed that participants assigned a significantly lower likability to both the female ($M = 3.25$, $SD = 0.89$, $p = 0.003$) and the nonbinary voice ($M = 3.11$, $SD = 0.82$, $p = 0.000$) than to the male voice ($M = 3.70$, $SD = 0.83$) (cf. Figure 5). An independent samples *t*-test further shows no significant differences ($p = 0.465$) in the likability of DVAs between participants who never used a DVA ($M = 3.23$, $SD = 0.95$) and those who have ($M = 3.31$, $SD = 0.83$), excluding a potential experience effect. Next, we examined the effect of the respondents' personal factors (i.e., respondents' gender, personality, egalitarianism, age and education) on the nonbinary voice's likability. For this, only nonbinary voice cases ($n = 161$) were considered.

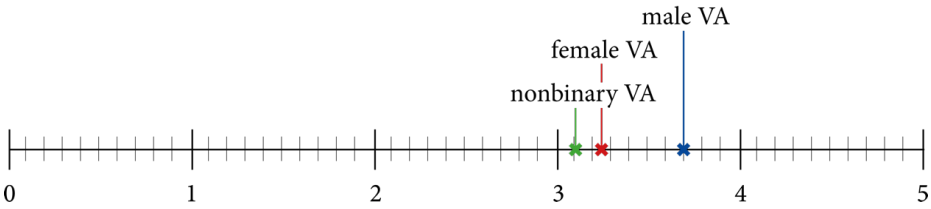



**Figure 5.** Likability of voices based on a 5-item semantic differential running from 1 = *dislike* to 5 = *like*.

#### 4.3.1. Respondents' Gender (H2a)

On average, women assigned a higher likability to the nonbinary voice option ($M = 3.23$, $SD = 0.82$) than men ($M = 2.96$, $SD = 0.81$) (cf. Figure 6). This difference between men and women is statistically significant: $t(159) = 2.0534$, $p = 0.042$. Although the effect size is small ($d = 0.3247$) [72], it provides support for our hypothesis H2a.

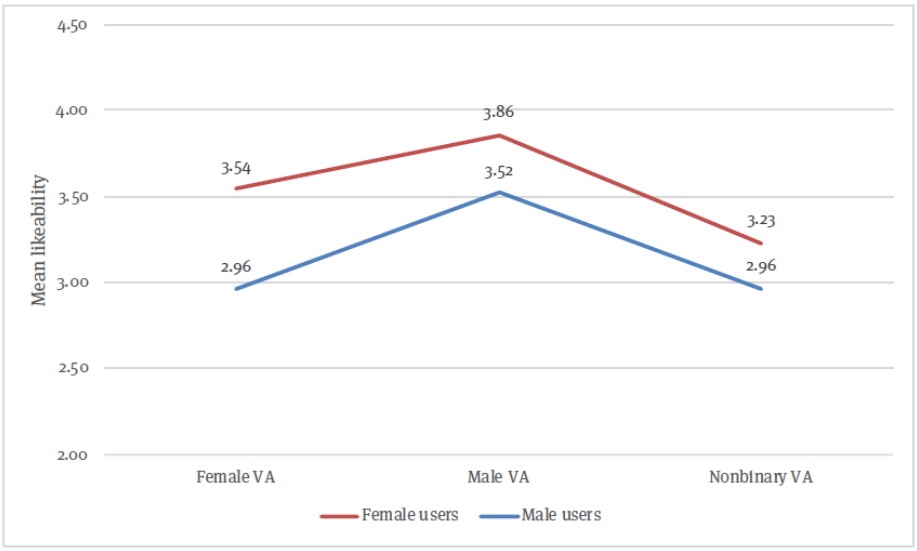



**Figure 6.** Likability of DVA voices by gender.

#### 4.3.2. Respondents' Personality (H2b)

An analysis of possible correlations between likability and the personality dimensions of the BFI-10 [68] revealed significance ($p < 0.01$) for agreeability ($r = 0.214$, $p = 0.006$) and neuroticism ($r = 0.214$, $p = 0.006$). These two dimensions are able to explain 9.65% of the variance found in our data on voice likability: $F(2, 158) = 8.438$, $p < 0.000$, $R2 = 0.0965$ (cf. Table 7). For the three other personality traits (i.e., conscientiousness, extraversion, and openness), the data does not reveal a significant connection with voice likability. Hence, we may argue that our hypothesis H2b is only partly supported, which opens up avenues for future investigations.

**Table 7.** Regression: neuroticism, agreeability, and likability.

| *Model Summary* | | | | | | |
|---|---|---|---|---|---|---|
| **Model** | **R** | **R Square** | **Adjusted R Square** | **Std. Error of the Estimate** | | |
| 1 | 0.311 [a] | 0.0965 | 0.0851 | 0.7885 | | |
| *ANOVA* [b] | | | | | | |
| **Model** | | **Sum of Squares** | **df** | **Mean Square** | **F** | **Sig.** |
| 1 | Regression | 10.4916 | 2 | 5.2458 | 8.4379 | 0.000 [c] |
| | Residual | 98.2279 | 158 | | | |
| | Total | 108.7195 | 160 | | | |
| *Coefficients* [d] | | | | | | |
| **Model** | | **Unstandardized Coefficients** | | | **t** | **Sig.** |
| | | **B** | **Std. Error** | | | |
| 1 | (Constant) | 1.8438 | 0.3182 | | 5.7940 | 0.000 |
| | Agreeability | 0.2329 | 0.0781 | | 2.9816 | 0.003 |
| | Neuroticism | 0.1908 | 0.0641 | | 2.9789 | 0.003 |

[a.] Predictors: (Constant), Neuroticism, Agreeability. [b.] Dependent Variable: LIKEABILITY. [c.] Predictors: (Constant), Neuroticism, Agreeability. [d.] Dependent Variable: LIKEABILITY.

#### 4.3.3. Respondents' Egalitarianism, Age and Education (H2c–e)

The eight-item GRSS included stereotypical masculine and feminine tasks. Both showed a good or acceptable internal consistency for masculine (Cronbach's $\alpha = 0.712$) and feminine tasks (Cronbach's $\alpha = 0.633$). However, a correlation analysis between GRSS and likability, based on the Spearman coefficient, was not significant ($p = 0.638$). Neither did an independent samples *t*-test show significant differences ($p = 0.623$) with respect to

likability of DVA voices between egalitarian-oriented people and those who tend to think in gender-stereotypical roles. Consequently, we had to reject our hypothesis H2c.

Based on the Spearman correlation analysis, which is appropriate for non-normal distributed data, we furthermore found that neither respondents' age ($p = 0.998$) nor their education ($p = 0.983$) exhibit a significant linear relationship with their likability of different DVA voices. Thus, we also had to reject our hypotheses H2d,e.

## 5. Discussion

Nass et al. [73]'s *Computers are Social Actors* (CASA) paradigm highlights the anthropomorphism of human characteristics and technology. Anthropological assumptions go hand-in-hand with the transfer of gender stereotypes [73,74]. As such, there is a tendency that people treat DVAs as they treat other people. The incorporation of gender cues into a DVA's design thus significantly affects human–machine interaction [36,74,75]. To promote inclusivity, properties of nonbinary DVAs can be utilized to mitigate or eschew these gender stereotypes.

### *5.1. Gender Perception in DVA Voices*

As society continues to break down the gender binary, more people become open about their self-perception of gender. Consequentially, the existence of nonbinary DVA voices, which do not implement distinct gender cues, also evolves [52]. Our pre-study results indicate that the nonbinary voice we chose for our investigation was suitable in terms of neutrality (male vs. female) and naturalness (computer-like vs. human-like). Still, only 24.8% of our study respondents perceived the voice as nonbinary, whereas 65.2% assigned it to a male gender. Although other studies match these percentages (e.g., [28,30]), it points to a still-missing awareness of non-gender-binary-conforming anthropomorphic technology.

To this end, some researchers uphold the notion that there is currently no truly gender-neutral voice [30,39]. Already in 1990, Butler and Trouble [8] discussed the idea of the gender spectrum. To represent gender's fluid concept, Schiebinger and Klinge [49] suggested the inclusion of nonbinary agents. However, to the broader society, the concept of the gender spectrum is a relatively recent topic. Therefore, users may still consider only binary options, which is why they seem to be inclined to assign the nonbinary DVA voice to either side of the spectrum. It is only with the rise of the LGBTQ+ community that gender fluid concepts become increasingly supported [76]. Thus, as the nonbinary gender option becomes less marginalized, users may start feeling more comfortable assigning a neutral gender to a disembodied agent. At the same time, limiting the gender of DVAs to three instead of two categories may in itself be counterproductive to mitigate gender stereotypes. Instead, we should aim for a better understanding of the cues that actually elicit gender stereotypes and learn how to control them so as to eventually create gender-neutral DVAs.

### *5.2. Elicitation of Gender Stereotypes*

Commercially available DVAs are often designed with a female persona in mind [77]. Moreover, it is commonly understood that humans transfer gender stereotypes to machines [32]. This becomes problematic when technology designers use these gender stereotypes to amplify existing societal norms and expectations. Anthropomorphism encourages users to rely on social norms and expectations to judge the fitness of DVAs. People are used to hearing female voices when interacting with DVAs and, consequently, connect them to descriptive and prescriptive gender-stereotypical beliefs. Therefore, the creation of gender-neutral or nonbinary DVAs may aid in eschewing such gender stereotypes [49]. That is, if users are undecided as to whether a DVA they interact with is male or female, they may not assign gender stereotypical traits to the agent, but rather a combination of different traits that are not bound to binary constraints.

Our study results indicate that gender-stereotypical traits are only elicited for the female DVA voice, whereas the male and nonbinary DVA voice did not show a significant correlation with stereotypical attribute assignment. It has to be noted, however, that

the perception of the nonbinary voice we used in this study leaned more toward male. Still, there was no statistically relevant difference between the assignment of feminine or masculine gender-stereotypical traits, confirming that nonbinary DVAs seem to not elicit gender-stereotypical trait attribution. As for our binary gender voice options (i.e., male and female), we have seen that respondents assign significantly more feminine than masculine trait attributes to the female DVA voice. For example, female vocal cues were connected with attributes such as *tender* and *gentle*, which can be clearly linked to gender stereotypes. These results are in accordance with previous studies [32,36], showing that DVA perception matches societal views [34]. Interestingly, the difference between male and female attributes assigned to the male DVA voice in our study was not significant. One cause for this disparity may be found in the vocal cues used for our voices, which may have evoked less gender-typical associations than voices used in other studies. For example, the $F_0$ fundamental frequency could have been too high in the male voice we selected for the study, which has been shown in previous work to affect masculinity ratings in listeners [78]. Moreover, our questionnaire setup may also have affected the assignment of trait attributes. That is, hearing DVA voice prompts instead of interacting with DVAs may elicit different traits.

In general, our study respondents rated feminine attributes higher than masculine attributes for all three genders. If, indeed, our perspectives are shaped by our experiences [79], it is not surprising that participants tend to hear feminine traits. Their experience is to hear females in the role as DVAs and, therefore, they transfer feminine traits to male and nonbinary voices equally. In other words, these biases are rooted in assumptions about what DVAs should sound like [80].

Considering the complexity of the topic, the prevalence of female-gender DVAs is not as simple as to say that users unanimously choose feminine voices. These preferences rather stem from gender-stereotypical assumptions shaped by cultural and social influences. Also, the innateness of industry (e.g., secretarial work) and their gender stereotypes consistently interfere with our potentially less stereotyped perception of DVAs [81].

### *5.3. Factors Influencing Likability*

While previous work predominately focused on the differences in the likability of gendered DVA voices (e.g., [10]), we concentrated on comparing a nonbinary voice option to the traditional male and female DVA voices. Although our male DVA voice was liked significantly more than the other two options, likability scores are above average for all three DVA voices. The preference for the male voice, and consequent lack of an expected higher likability for the female voice, may have been caused by certain voice characteristics. As already outlined earlier, the male voice also did not elicit male stereotypes at the level that we had expected, and so we suspect that these characteristics also affected the voices likability, putting it ahead of the female voice in the comparative ranking.

Furthermore, it seems interesting that, although personal factors have been shown to affect the likability of gendered DVA voices, and may thus equally apply to nonbinary DVA voices, our study results do not point to any such connections between voice likability and respondents' age, education, or employment status. In terms of respondents' gender, our data support neither the similarity attraction effect [22,59], nor the cross-gendered effect [82]. The only statistically significant result we found is that women in general rated all three gender options higher in likability than men did. A previous work by Cambre et al. [83] comparing 18 different text-to-speech voices did not find such a significant difference in how voices are perceived by different genders. Thus, it seems the consistently higher ratings by our female participants may have been the result of the distinct voice we have selected for this study.

Yet, two dimensions of respondents' personality, i.e., their agreeability and neuroticism, show a positive correlation with the DVA voice likability. Agreeability, for example, manifests itself in behavioral characteristics perceived as kind and sympathetic. People who score high on neuroticism, on the other hand, are more likely than average to be

moody [84]. Nonbinary voices can be assigned similar traits, such as sympathetic, gentle, or even moody [15].

## 6. Conclusions, Limitations and Future Outlook

Whereas the focus of prior work has been primarily on investigating the connection between binary DVA voices and gender stereotypes, we investigated the ascription of gender stereotypes and the effect of personal factors on the nonbinary voice option and its likability. To this end, our investigation yielded three main contributions to the theory of gender-studies in human–machine interaction: (1) Our results indicate poorer likability for both our nonbinary and female DVA voices, especially compared to the male DVA voice; (2) we found that women express higher likability for all our DVA voices than men; and (3) in terms of personality, we found that neuroticism and agreeability correlate with the likability of our nonbinary voice option.

### *6.1. Limitations and Areas for Future Work*

While our study contributes to a more comprehensive understanding of nonbinary DVA voices and their manifestation of gender stereotypes, several limitations must be acknowledged. Apart from common limitations such as a lack of diversity found in the sample, or too socially desirable and biased responses, we particularly want to address limitations of the overall research design and make respective suggestions for further research.

Our research design was limited in that study respondents were simply listening to DVA voice prompts. Yet, some trait attributions may only be evoked when interacting with the DVA in an experimental setting, or over a more extended period. Thus, future work should focus on how more extended interactions with a nonbinary DVA affect the perception and the elicitation of gender stereotypes. Furthermore, creating a nonbinary voice is highly complex, and its characteristics, such as tonality, pitch and choice of words, affect a user's perception. As such, our study does not provide a holistic understanding of nonbinary DVA voices. Consequently, a deeper investigation of nonbinary voice characteristics and their effects on user perception is needed. Moreover, although the results show that our nonbinary DVA voice scored the lowest out of all three voice options regarding its likability, we did not collect any data on people's preferences; i.e., they were randomly assigned to one voice, and thus did not have the option to select or express their preference for a male, female, or nonbinary voice. Further research should thus investigate people's voice preferences in more detail, and potentially also explore its link to technology acceptance.

Also, our approach to understanding user's personality and its effect on nonbinary DVA voices requires expansion and deeper investigation. In particular, the vague correlation of personality factors needs to be further explored. To this end, one may even consider exploring connections between political views and the perception of gender-neutrality, as well as the perceptions that members of the LGBTQ+ community have towards nonbinary voices, particularly since they represent another group that has been subject to stereotypical discrimination and ignorance. Eventually, we may then be able to build a dedicated machine learning model that is capable of predicting people's voice preferences, and make respective adjustments during real-time DVA interactions.

Finally, regarding gender-stereotypical attributes, we acknowledge that six attributes per gender may not be sufficient to comprehensively illustrate gender stereotypes. These stereotypes are deeply ingrained in people, and are highly complex to map. While we feel that it is essential to question the reasons for primarily female-gendered DVAs, we raise the issue of whether a gender-neutral DVA is actually able to change gender norms and expectations. In particular, there is the potential of disquieting users, which might ultimately result in decreasing user satisfaction. Society's norms and expectations are continually evolving, and it is therefore vital that future research aims to investigate the long-term changes in gender bias connected to nonbinary DVAs.

### *6.2. Concluding Remarks*

Technology companies offer DVAs with female-gendered voices as default and, as such, female DVA voices prevail in most commercially used systems. At the same time, female-gendered DVAs are closely tied to their gender-stereotypical trait attribution. The perception of women, even artificial ones, in assistant or servitude positions reflects gender bias and harmful culture. Our work shows that nonbinary DVA voices do not elicit such gender-stereotypical trait attribution and, as such, that their use disrupts the reinforcement of negative stereotypes. We thus propose that, by designing DVAs and their voices nonbinary (i.e., in a way that does not amplify gender stereotypes), we achieve a mitigation of gender bias and may thus promote gender equality in anthropomorphic technology. Furthermore, offering voices that are not constrained by a binary gender construct may be seen as an opportunity to create systems that reflect society and to actively promote the inclusion of diverse people. Rather than becoming yet another weapon to amplify gender-stereotypical biases by re-creating them in technological artifacts, DVAs can, therefore, become the means to drive change and progress gender equality.

**Author Contributions:** The article was a collaborative effort by all co-authors. Conceptualization, S.T.L., S.S. and W.S.M.T.v.K.; methodology, S.T.L.; validation, S.T.L., S.S. and W.S.M.T.v.K.; formal analysis, S.T.L.; investigation, S.T.L.; resources, S.T.L. and S.S.; data curation, S.T.L. and S.S.; writing—original draft preparation, S.T.L., S.S. and A.E.; writing—review and editing, S.S., W.S.M.T.v.K. and T.S.; visualization, S.T.L. and S.S.; supervision, S.S.; project administration, S.T.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of MCI—The Entrepreneurial School.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The data presented in this study are openly available via Zenodo at https://doi.org/10.5281/zenodo.11468262 (accessed on 4 June 2024).

**Acknowledgments:** We want to thank the Acapela Group (available online: https://www.acapela-group.com/ [accessed on 4 June 2024]) for the production of the nonbinary synthetic voice which we used in our study.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| DVA | Digital Voice Assistant |
| BSRI | Bem Sex Role Inventory |
| BFI | Big Five Inventory |
| GRSS | Gender Role Stereotype Scale |
| ATI | Affinity for Technology Interaction |
| SFF | Speaking Fundamental Frequency |
| ANOVA | Analysis of Variance |
| CASA | Computers are Social Actors |

## References

1. Synup Corporation. 80+ Industry Specific Voice Search Statistics for 2024. Available online: https://www.synup.com/voice-search-statistics (accessed on 8 March 2024).
2. Semrush Blog. 7 Up-to-Date Voice Search Statistics (+3 Best Practices). Available online: https://www.semrush.com/blog/voice-search-statistics/ (accessed on 8 March 2024).
3. Yaguara. 79+ Voice Search Statistics for 2024 (Data, Users & Trends). Available online: https://www.yaguara.co/voice-search-statistics/ (accessed on 8 March 2024).
4. Serpwatch. Voice Search Statistics: Smart Speakers, Voice Assistants, and Users in 2024. Available online: https://serpwatch.io/blog/voice-search-statistics/ (accessed on 8 March 2024).

5. UNESCO. I'd Blush If I Could: Closing Gender Divides in Digital Skills through Education. Available online: https://unesdoc.unesco.org/ark:/48223/pf0000367416.locale=en/ (accessed on 8 March 2024).
6. Nass, C.; Moon, Y. Machines and mindlessness: Social responses to computers. *J. Soc. Issues* **2000**, *56*, 81–103. [CrossRef]
7. Otterbacher, J.; Talias, M. S/he's too Warm/Agentic! The Influence of Gender on Uncanny Reactions to Robots. In Proceedings of the 2017 12th ACM/IEEE International Conference on Human-Robot Interaction, Vienna, Austria, 6–9 March 2017; pp. 214–223.
8. Butler, J.; Trouble, G. Feminism and the Subversion of Identity. *Gend. Troubl.* **1990**, *3*, 1–25.
9. McTear, M.F.; Callejas, Z.; Griol, D. *The Conversational Interface—Talking to Smart Devices*; Springer: Berlin/Heidelberg, Germany, 2016; Volume 6.
10. Nass, C.I.; Brave, S. *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship*; MIT Press: Cambridge, MA, USA, 2005.
11. Nass, C.; Gong, L. Speech interfaces from an evolutionary perspective. *Commun. ACM* **2000**, *43*, 36–43. [CrossRef]
12. McTear, M.F. Spoken dialogue technology: Enabling the conversational user interface. *ACM Comput. Surv.* **2002**, *34*, 90–169. [CrossRef]
13. Politt, R.; Pollock, J.; Waller, E. *Day-to-Day Dyslexia in the Classroom*; Routledge: Oxfordshire, UK, 2004.
14. Kiss, G. Autonomous agents, AI and chaos theory. In Proceedings of the First International Conference on Simulation of Adaptive Behavior (From Animals to Animats), Paris, France, 14 February 1991; Citeseer: Princeton, NJ, USA, 1991.
15. Braun, M.; Mainz, A.; Chadowitz, R.; Pfleging, B.; Alt, F. At your service: Designing voice assistant personalities to improve automotive user interfaces. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, Scotland, UK, 4–9 May 2019; pp. 1–11.
16. Wooldridge, M.; Jennings, N.R. Intelligent agents: Theory and practice. *Knowl. Eng. Rev.* **1995**, *10*, 115–152. [CrossRef]
17. Hamill, L. Controlling smart devices in the home. *Inf. Soc.* **2006**, *22*, 241–249. [CrossRef]
18. Gaiani, M.; Benedetti, B. A methodological proposal for representation and scientific description of the great archaeological monuments. In Proceedings of the 2014 International Conference on Virtual Systems & Multimedia (VSMM), Hong Kong, China, 9–12 December 2014; pp. 122–129.
19. Seeger, A.M.; Pfeiffer, J.; Heinzl, A. When do we need a human? Anthropomorphic design and trustworthiness of conversational agents. In Proceedings of the ACM CHI onference on Human Factors in Computing Systems, Denver, CO, USA, 6–11 May 2017.
20. Wagner, K.; Nimmermann, F.; Schramm-Klein, H. Is it human? The role of anthropomorphism as a driver for the successful acceptance of digital voice assistants. In Proceedings of the 52nd Hawaii International Conference on System Sciences, Maui, HI, USA, 8–11 January 2019.
21. Hoy, M.B. Alexa, Siri, Cortana, and more: An introduction to voice assistants. *Med Ref. Serv. Q.* **2018**, *37*, 81–88. [CrossRef] [PubMed]
22. Cambre, J.; Kulkarni, C. One Voice Fits All? Social Implications and Research Challenges of Designing Voices for Smart Devices. *Proc. Acm Hum. Comput. Interact.* **2019**, *3*, 1–19. [CrossRef]
23. Byrne, D.; Nelson, D. The effect of topic importance and attitude similarity-dissimilarity on attraction in a multistranger design. *Psychon. Sci.* **1965**, *3*, 449–450. [CrossRef]
24. Zuckerman, M.; Driver, R.E. What sounds beautiful is good: The vocal attractiveness stereotype. *J. Nonverbal Behav.* **1989**, *13*, 67–82. [CrossRef]
25. Kramer, J.; Noronha, S.; Vergo, J. A user-centered design approach to personalization. *Commun. ACM* **2000**, *43*, 44–48. [CrossRef]
26. Cole, E.R. Intersectionality and research in psychology. *Am. Psychol.* **2009**, *64*, 170. [CrossRef] [PubMed]
27. Haslanger, S. Ontology and social construction. *Philos. Top.* **1995**, *23*, 95–125. [CrossRef]
28. Piper, A.M. Stereotyping Femininity in Disembodied Virtual Assistants. Master's Thesis, Iowa State University, Ames, Iowa, 2016.
29. Hyde, J.S.; Bigler, R.S.; Joel, D.; Tate, C.C.; van Anders, S.M. The future of sex and gender in psychology: Five challenges to the gender binary. *Am. Psychol.* **2019**, *74*, 171. [CrossRef]
30. Bryant, D.; Borenstein, J.; Howard, A. 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, Cambridge, UK, 23–26 March 2020; pp. 13–21.
31. Søraa, R.A. Mechanical genders: How do humans gender robots? *Gender, Technol. Dev.* **2017**, *21*, 99–115. [CrossRef]
32. Nass, C.; Moon, Y.; Green, N. Are machines gender neutral? Gender-stereotypic responses to computers with voices. *J. Appl. Soc. Psychol.* **1997**, *27*, 864–876. [CrossRef]
33. Nomura, T. Robots and gender. *Gend. Genome* **2017**, *1*, 18–25. [CrossRef]
34. Prentice, D.A.; Carranza, E. What women and men should be, shouldn't be, are allowed to be, and don't have to be: The contents of prescriptive gender stereotypes. *Psychol. Women Q.* **2002**, *26*, 269–281. [CrossRef]
35. Brahnam, S.; De Angeli, A. Gender affordances of conversational agents. *Interact. Comput.* **2012**, *24*, 139–153. [CrossRef]
36. Eyssel, F.; Kuchenbrandt, D.; Bobinger, S.; De Ruiter, L.; Hegel, F. 'If you sound like me, you must be more human': On the interplay of robot and user features on human-robot acceptance and anthropomorphism. In Proceedings of the HRI'12—Proceedings of the 7th Annual ACM/IEEE International Conference on Human-Robot Interaction, Boston, MA, USA, 5–8 March 2012; pp. 125–126. [CrossRef]

37. Adams, R.; Loideáin, N.N. Addressing indirect discrimination and gender stereotypes in AI virtual personal assistants: The role of international human rights law. *Camb. Int. Law J.* **2019**, *8*, 241–257. [CrossRef]
38. Bergen, H. 'I'd blush if I could': Digital assistants, disembodied cyborgs and the problem of gender. *Word Text, J. Lit. Stud. Linguist.* **2016**, *6*, 95–113.
39. Oudshoorn, N.; Rommes, E.; Stienstra, M. Configuring the User as Everybody: Gender and Design Cultures in Information and Communication Technologies. *Sci. Technol. Hum. Values* **2004**, *29*, 30–63. [CrossRef]
40. Balsamo, A.M. *Technologies of the Gendered Body: Reading Cyborg Women*; Duke University Press: Durham, NC, USA, 1996.
41. Venkatesh, V.; Morris, M.G.; Ackerman, P.L. A longitudinal field investigation of gender differences in individual technology adoption decision-making processes. *Organ. Behav. Hum. Decis. Process.* **2000**, *83*, 33–60. [CrossRef]
42. Crowell, C.R.; Scheutz, M.; Schermerhorn, P.; Villano, M. Gendered voice and robot entities: Perceptions and reactions of male and female subjects. In Proceedings of the 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2009, St. Louis, MO, USA, 10–15 October 2009; pp. 3735–3741. [CrossRef]
43. Schermerhorn, P.; Scheutz, M.; Crowell, C.R. Robot social presence and gender: Do females view robots differently than males? In Proceedings of the HRI 2008—Proceedings of the 3rd ACM/IEEE International Conference on Human-Robot Interaction: Living with Robots, Amsterdam, The Netherlands, 12–15 March 2008; pp. 263–270. [CrossRef]
44. Kuo, I.H.; Rabindran, J.M.; Broadbent, E.; Lee, Y.I.; Kerse, N.; Stafford, R.M.Q.; MacDonald, B.A. Age and gender factors in user acceptance of healthcare robots. In Proceedings of the 18th IEEE International Symposium on Robot and Human Interactive Communication, Toyama, Japan, 27 September–2 October 2009; pp. 214–219.
45. Wang, Y.; Young, J.E. Beyond "pink" and "blue": Gendered attitudes towards robots in society. In *Gender and IT Appropriation. Science and Practice on Dialogue—Forum for Interdisciplinary Exchange*; European Society for Socially Embedded Technologies: Siegen, Germany, 2014; pp. 49–59.
46. Siegel, M.; Breazeal, C.; Norton, M.I. Persuasive robotics: The influence of robot gender on human behavior. In Proceedings of the 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2009, St. Louis, MO, USA, 11–15 October 2009; pp. 2563–2568. [CrossRef]
47. Rhim, J.; Kim, Y.; Kim, M.S.; Yim, D.Y. The effect of gender cue alterations of robot to match task attributes on user's acceptance perception. In Proceedings of the HCI Korea 2015, Seoul, Republic of Korea, 10–12 December 2014; pp. 51–57.
48. Curry, A.C.; Rieser, V. # MeToo Alexa: How conversational systems respond to sexual harassment. In Proceedings of the Second Acl Workshop on Ethics in Natural Language Processing, New Orleans, LA, USA, 5 June 2018 ; pp. 7–14.
49. Schiebinger, L.; Klinge, I. Gendered innovations. In *How Gender Analysis Contributes to Research*; Publications Office of the European Union, Directorate General for Research & Innovation: Brussels, Belgium, 2013.
50. Søndergaard, M.L.J.; Hansen, L.K. Intimate Futures: Staying with the Trouble of Digital Personal Assistants through Design Fiction. In Proceedings of the 2018 Designing Interactive Systems Conference, Hong Kong, China, 9–13 June 2018; pp. 869–880.
51. Phan, T. The Materiality of the Digital and the Gendered Voice of Siri. *Transformations* **2017**, *29*, 24–33.
52. Davies, S.; Papp, V.G.; Antoni, C. Voice and communication change for gender nonconforming individuals: Giving voice to the person inside. *Int. J. Transgenderism* **2015**, *16*, 117–159. [CrossRef]
53. Schmid, M.; Bradley, E. Vocal pitch and intonation characteristics of those who are gender non-binary. In Proceedings of the 19th International Conference of Phonetic Sciences, Melbourne, Australia, 5–9 August 2019; pp. 2685–2689.
54. Stoicheff, M.L. Speaking fundamental frequency characteristics of nonsmoking female adults. *J. Speech Lang. Hear. Res.* **1981**, *24*, 437–441. [CrossRef] [PubMed]
55. Titze, I. *Principles of Voice Production*; Prentice Hall: Hoboken, NJ, USA, 1994.
56. Zheng, J.F.; Jarvenpaa, S. Negative Consequences of Anthropomorphized Technology: A Bias-Threat-Illusion Model. In Proceedings of the 52nd Hawaii International Conference on System Sciences, Maui, HI, USA, 8–11 January 2019.
57. Jørgensen, S.H.; Baird, A.; Juutilainen, F.T.; Pelt, M.; Højholdt, N.C. [multi'vocal]: Reflections on engaging everyday people in the development of a collective non-binary synthesized voice. In Proceedings of the EVA Copenhagen 2018, Aalborg University, Copenhagen, Denmark, 15–17 May 2018. [CrossRef]
58. Turner, J.C.; Oakes, P.J. The significance of the social identity concept for social psychology with reference to individualism, interactionism and social influence. *Br. J. Soc. Psychol.* **1986**, *25*, 237–252. [CrossRef]
59. Lee, E.J.; Nass, C.; Brave, S. Can computer-generated speech have gender? An experimental test of gender stereotype. In Proceedings of the CHI'00 extended abstracts on Human factors in computing systems, The Hague, The Netherlands, 1–6 April 2000; pp. 289–290.
60. Strait, M.; Briggs, P.; Scheutz, M. Gender, more so than age, modulates positive perceptions of language-based human-robot interactions. In Proceedings of the AISB Convention, Canterbury, UK, 20–22 April 2015.
61. Nass, C.; Moon, Y.; Fogg, B.J.; Reeves, B.; Dryer, C. Can computer personalities be human personalities? In Proceedings of the Conference Companion on Human Factors in Computing Systems, Denver, CO, USA, 7–11 May 1995; pp. 228–229.
62. Moskowitz, G.B.; Li, P. Egalitarian goals trigger stereotype inhibition: A proactive form of stereotype control. *J. Exp. Soc. Psychol.* **2011**, *47*, 103–116. [CrossRef]
63. Chang, R.C.S.; Lu, H.P.; Yang, P. Stereotypes or golden rules? Exploring likable voice traits of social robots as active aging companions for tech-savvy baby boomers in Taiwan. *Comput. Hum. Behav.* **2018**, *84*, 194–210. [CrossRef]

64. Nomura, T.; Takagi, S. Exploring effects of educational backgrounds and gender in human-robot interaction. In Proceedings of the 2011 International conference on user science and engineering (i-user), Selangor, Malaysia, 29 November–1 December 2011; pp. 24–29.
65. Monahan, J.L. I don't know it but I like you: The influence of nonconscious affect on person perception. *Hum. Commun. Res.* **1998**, *24*, 480–500. [CrossRef]
66. Bartneck, C.; Kulić, D.; Croft, E.; Zoghbi, S. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *Int. J. Soc. Robot.* **2009**, *1*, 71–81. [CrossRef]
67. Bem, S.L. The measurement of psychological androgyny. *J. Consult. Clin. Psychol.* **1974**, *42*, 155. [CrossRef] [PubMed]
68. Rammstedt, B.; John, O.P. Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *J. Res. Personal.* **2007**, *41*, 203–212. [CrossRef]
69. Mills, M.J.; Culbertson, S.S.; Huffman, A.H.; Connell, A.R. Assessing gender biases: Development and initial validation of the gender role stereotypes scale. *Gend. Manag. Int. J.* **2012**, *27*, 520–540. [CrossRef]
70. Attig, C.; Wessel, D.; Franke, T. Assessing personality differences in humantechnology interaction: An overview of key self-report scales to predict successful interaction. In Proceedings of the International Conference on Human-Computer Interaction, Vancouver, BC, Canada, 9–14 July 2017; pp. 19–29.
71. George, D.; Mallery, P. *SPSS for Windows Step by Step: A Simple Guide and Reference, 17.0 Update*; Allyn & Bacon: Boston, MA, USA, 2010.
72. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*; Academic Press: Cambridge, MA, USA, 2013.
73. Nass, C.; Steuer, J.; Tauber, E.R. Computers are social actors. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Boston, MA, USA, 24–28 April 1994; pp. 72–78.
74. Tay, B.; Jung, Y.; Park, T. When stereotypes meet robots: The double-edge sword of robot gender and personality in human–robot interaction. *Comput. Hum. Behav.* **2014**, *38*, 75–84. [CrossRef]
75. Powers, A.; Kramer, A.D.; Lim, S.; Kuo, J.; Lee, S.L.; Kiesler, S. Eliciting information from people with a gendered humanoid robot. In Proceedings of the IEEE International Workshop on Robot and Human Interactive Communication, Nashville, TN, USA, 13–15 August 2005; Volume 2005, pp. 158–163. [CrossRef]
76. Danielescu, A. Eschewing Gender Stereotypes in Voice Assistants to Promote Inclusion. In Proceedings of the CUI '20: Proceedings of the 2nd Conference on Conversational User Interfaces, Bilbao, Spain, 22–24 July 2020; ACM Association for Computing Machinery; ACM: New York, NY, USA, 2020; pp. 1–3.
77. Habler, F.; Schwind, V.; Henze, N. Effects of smart virtual assistants' gender and language. In Proceedings of the Mensch und Computer 2019, ACM International Conference Proceeding Series. ACM Association for Computing Machinery, Hamburg, Germany, 8–11 September 2019; pp. 469–473.
78. Cartei, V.; Bond, R.; Reby, D. What makes a voice masculine: Physiological and acoustical correlates of women's ratings of men's vocal masculinity. *Horm. Behav.* **2014**, *66*, 569–576. [CrossRef] [PubMed]
79. Kerr, A.D. Alexa and the Promotion of Oppression. In Proceedings of the 2018 ACM Celebration of Women in Computing (womENcourage'18), Belgrade, Serbia, 3–5 October 2018; ACM: New York, NY, USA, 2018.
80. Heilman, M.E. Gender stereotypes and workplace bias. *Res. Organ. Behav.* **2012**, *32*, 113–135. [CrossRef]
81. Gaucher, D.; Friesen, J.; Kay, A.C. Evidence That Gendered Wording in Job Advertisements Exists and Sustains Gender Inequality. *J. Personal. Soc. Psychol.* **2011**, *101*, 109–128. [CrossRef]
82. Alexander, E.; Bank, C.; Yang, J.J.; Hayes, B.; Scassellati, B. Asking for Help from a Gendered Robot. In Proceedings of the Annual Meeting of the Cognitive Science Society, Quebec City, QC, Canada, 23–26 July 2014.
83. Cambre, J.; Colnago, J.; Maddock, J.; Tsai, J.; Kaye, J. Choice of Voices: A Large-Scale Evaluation of Text-to-Speech Voice Quality for Long-Form Content. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 25–30 April 2020; ACM: New York, NY, USA, 2020; CHI '20, pp. 1–13. [CrossRef]
84. Costa, P.T.; McCrae, R.R. The five-factor model of personality and its relevance to personality disorders. *Sci. Ment. Heal. Vol. Personal. Personal. Disord.* **1992**, *6*, 17–33. [CrossRef]


**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.