Aug 10th, 12:00 AM

The Impact of Gender Stereotyping on the Perceived Likability of Virtual Assistants

Claus-Peter H. Ernst
CBS International Business School, cernst@fb3.fra-uas.de

Nils Herm-Stapelberg
Johannes Gutenberg-Universität Mainz, hermstapelberg@uni-mainz.de

The Impact of Gender Stereotyping on the Perceived Likability of Virtual Assistants

Completed Research

Claus-Peter H. Ernst
CBS International Business School
c.ernst@cbs.de

Nils Herm-Stapelberg
Johannes Gutenberg-Universität Mainz
hermstapelberg@uni-mainz.de

Abstract

Research on gender stereotyping suggests that females are perceived as being more likable than males are. Based on the CASA paradigm, which describes the human tendency to assign human traits to computers, we expect that this stereotype might also be present for virtual assistants, i.e., that female-voice virtual assistants are perceived as being more likable than male-voice virtual assistants. We conduct a controlled experiment that simulates a lifelike interaction with differently voiced virtual assistants to test this hypothesis. The results emphasize that gender stereotypes indeed influence users' perception of virtual assistants, where female-voiced assistants are perceived as being more likable than male-voiced assistants. This has practical implications for the development and design of devices that utilize these virtual assistants.

Keywords

Virtual assistant, gender stereotyping, likability.

Introduction

Virtual assistants such as Alexa and Siri are on the rise. Present on smartphones for a number of years, they are able to perform tasks such as reminding the user about an upcoming flight, simply by interacting with the user through voice. Multiple companies have successfully launched devices that contain virtual assistants in the context of the smart home, such as Amazon's smart-speaker line Echo.

According to the CASA paradigm (Nass et al. 1994), people are known to assign human traits to computers. In line with this, Ernst and Herm-Stapelberg (2020) showed that gender stereotypes, which are "psychological traits … that are believed to occur with differential frequency in …" males and females (Best 2004, p. 11), can also arise when using virtual assistants. More specifically, the authors built on the corresponding gender stereotype and showed that virtual assistants that are perceived as being male are perceived to be more competent than virtual assistants perceived as being female, emphasizing that technology companies should be aware of gender stereotypes in general.

While competence is an important utilitarian-based trait of virtual assistants, hedonic motivations also influence the usage and adoption of information systems and technologies in general (e.g., Davis et al. 1992). Indeed, likability, that is, "the degree to which … [a] person is perceived as friendly, nice, polite and pleasant to be around" (Pulles and Hartman 2017, p. 57), is a hedonic-based trait that is also widely accepted as an important quality to have in order to be successful at a job (cf. Schneider et al. 2010). Hence, we believe that being perceived as likable, and not only as competent, is crucial for maintaining the strong growth trajectory of virtual assistants and their connected devices, such as smart speakers.

Research on gender stereotyping suggests that females are associated with adjectives that carry the notion of likability such as "pleasant" more frequently than males are (e.g., Williams and Best 1990).[1] We thus

[1] In order to address any misunderstanding, we would like to emphasize that this does not mean that males are in fact less likeable than females. In fact, we do not believe that gender determines likability. Rather, in this study, we seek to build a deeper understanding on how gender stereotyping may influence our everyday life.

seek to build on the study of Ernst and Herm-Stapelberg (2020) and contribute to the following research question by drawing from the CASA paradigm: Does gender stereotyping influence the perception of virtual assistants with regard to their perceived likability?

In order to test our hypothesis, we assigned 49 participants to two groups and asked them to interact with a virtual assistant by providing a predefined list of eight different requests. While one group's virtual assistant answered with a female voice, the other group's virtual assistant answered with a male voice. We conducted a survey after the interaction and asked the participants to indicate how likable they perceived the virtual assistant to be. Finally, by comparing the likability levels perceived by both groups we found that female-voiced virtual assistants are indeed perceived to be more likable than male-voiced virtual assistants.

The paper is structured as follows: First, we will introduce the notion of gender stereotyping as well as the CASA paradigm. Next, we will build our hypotheses and describe our research design as well as show the results of our experiment. We will conclude our article by describing the limitations and implications of our empirical study.

Theoretical Background

Gender Stereotyping

Multiple studies have confirmed that gender stereotypes, i.e., psychological characteristics that are commonly assigned to a specific gender, are present across and within different cultures (Williams et al. 1999, p. 513). In a large-scale cross-cultural study by Williams and Best (1990), university students had to rate whether each one of 300 adjectives of the Adjective Check List (Gough and Heilbrun 1983) was more often associated with a certain gender. The authors were able to show that there is a "high degree of pancultural similarity in the patterns of characteristics differentially associated with women and men in the 25 countries studied" (Williams et al. 1999, p. 514), indicating the presence of gender stereotypes. More specifically, females were associated with adjectives conveying likability (e.g., "pleasant") as well as with adjectives that imply being supporting and nurturing. Men however were described as strong, dominant, or autonomous.

Williams et al. (1999) further analyzed the data of Williams and Best (1990) and found that characteristics such as extraversion, emotional stability and openness to experience were more strongly associated with men, whereas agreeableness was more strongly associated with women.

A similar study carried out in the US was conducted by Sherriffs and McKee (1957) who let respondents assign 200 adjectives to either males or females. They showed that adjectives associated with the stereotype of women can be clustered into three categories, i.e. "social skills and grace", "warmth and emotional support" as well as "spiritual implications of experience" (Sherriffs and McKee 1957, pp. 453-454). The male stereotype was also associated with three clusters, namely a "straightforward uninhibited social style", "rational competence and ability" as well as "action and effectiveness". (Sherriffs and McKee 1957, p. 452).

Computers as Social Actors

The "computers are social actors" (CASA) paradigm describes the human tendency to unconsciously assign typical human traits, such as intelligence, to machines during their interactions with them (Nass and Lee 2000; Nass et al. 1994). Users treat the inanimate objects as if they were interacting with an actual person, and have the feeling that the machine can assert dominance (Nass et al. 1997) or be extroverted (Nass and Lee 2000).

A trait that is often assigned to machines is the notion of gender. Multiple studies have shown that humanlike features, such as voice, facial cues, or long hair, have an influence on the perceived gender of the robot or computer (Edwards et al. 2019; Eyssel and Hegel 2012; Nass et al. 1997). More precisely, traits that are commonly associated with female traits, such as longer hair or higher pitched voices, lead to the perception that the machine is actually female.

Since users can perceive a computer as male or female, it is interesting to ask whether gender stereotypes influence the perception of users' interactions with these computers. Studies show that users prefer "gendered" robots when the assigned task matches gender stereotypes (Carpenter et al. 2009; Eyssel and Hegel 2012; Tay et al. 2014). More specifically, in the context of health care, users prefer robots perceived as female, whereas in the context of security, users prefer robots perceived as male, despite the fact that both robots have the same abilities. Nass et al. (1997) also showed that vocal cues are sufficient for the assignment of gender and gender stereotypes, even when users are specifically told that the interaction is not with a person but with a genderless computer, and even if there are no other indications of gender, such as facial cues.

Even though the CASA paradigm has been extensively applied in multiple areas, there are only few studies concerning the use of smart home automation, for example, Siri or Alexa. Damen and Toh (2019) examined how the gender of automated agents influences the trust users have in them. Similar to Tay et al. (2014), they found that virtual assistants are trusted more when their perceived gender fits the task environment. More precisely, female voiced assistants are preferred in stereotypically female environments, such as at home, and male voice assistants are preferred in stereotypically male environments, such as at the office. Ernst and Herm-Stapelberg (2020) examined the perceived competence of male- and female-voiced personal assistants, based on classic gender stereotypes, and indeed found that male voiced assistants are perceived as being more competent than their female counterparts.

As mentioned, competence as a utilitarian motive is not the only interesting trait to explain why users adopt or reject using personal assistants. However, so far, it is largely unclear how gender stereotypes influence the hedonic dimensions of virtual assistant use, such as the likability of a system.

Research Model

The CASA paradigm, which has been confirmed in numerous studies, postulates that people assign human traits to computers when interacting with them (Nass et al. 1994). Additionally, people assign specific genders to a computer, for example, based on vocal cues (Nass et al. 1997). Drawing from these two central findings, it has also been shown that gender stereotypes – behaviors and traits that are assigned differently to males and females (e.g., Williams et al. 1999) – can also be present for virtual assistants (Ernst and Herm-Stapelberg 2020). In summary, if the voices of virtual assistants are designed correspondingly (e.g., by using a higher or lower pitched voice (Eyssel and Hegel 2012; Eyssel et al. 2011)), people perceive them as being either male or female and may thus assign different traits and characteristics to them.

Among other factors, hedonic motivations are known to drive technology usage (e.g., Davis et al. 1992). As a result, likability, as a hedonic-based trait, is likely one of the most important qualities for virtual assistants to have. In the context of gender stereotyping, research suggests that females are more frequently associated with the notion of likability than males are (e.g., Williams and Best 1990). Drawing from the CASA paradigm, we expect this perception of likability to also be present in the context of gendered virtual assistants. More specifically, we hypothesize the following: *Virtual assistants that are perceived as being female will be perceived as being more likable than those that are perceived as being male.*

Research Design

Experiment

To test this hypothesis, we built an experiment using a between-subjects design[2]. More precisely, we provided all of our participants with a list of eight requests, which represent common tasks that virtual

[2] In our context, results from a within-subject design would have been severely flawed, since the respondents would have easily been able to discover the purpose of the study (i.e., the different voices of the virtual assistants) and, thus, sponsorship effects, sequence effects, and memory effects would come up.

assistant can fulfill (Ernst and Herm-Stapelberg 2020). We told the participants that we placed a smart speaker inside a non-transparent box and asked them to make its included virtual assistant perform these tasks. Whereas one group's speaker answered with a female voice, the other group's speaker answered with a male voice. However, we had not actually placed a real smart speaker inside the box, and instead had placed a Bluetooth speaker there. After each request, we played a corresponding prerecorded answer[3] through the Bluetooth speaker using an iPhone in order to create the illusion of a smart speaker for the participants. We were thus able to make sure that all participants would receive the exact same answers, to avoid issues with misunderstood language (e.g., because of dialect) or any brand-based bias. Since virtual assistants are often not able to give a perfect answer, misunderstand a request, or may be programmed to vary the exact used words when answering, some answers were formulated in an imprecise way (3, 6), one was answered incorrectly (2), and one was not answered at all (4) — thus providing a more realistic experience. Table 1 presents the translated list of provided requests and transcriptions of the prerecorded answers, since the study was conducted in the German language. We also used "Computer" as the wake-up word in order to further avoid any gender-based bias that might occur due to a male or female name (e.g., Alexa).

No.	Computer, …	Answer
1	What day is it?	It is Tuesday, June 11th 2019.
2	How many milliliters is 30 centiliters?	30 millimeters are 3 centimeters.
3	What is 30 percent of 69 €?	30 percent of 69 is 20.7.
4	How far is Mainz from Berlin?	I am sorry, I cannot help you with that.
5	How old is Barack Obama?	Barack Obama is 57 years old.
6	How many days is it until Christmas?	It is 6 months until Christmas.
7	Flip a coin.	It shows heads.
8	Is it going to rain tomorrow?	It does not look like it's going to rain tomorrow.

Table 1. Virtual Assistant Requests and Answers

After the experiment, we conducted a questionnaire in which the respondents were asked to rate how likable they found the virtual assistant to be. Six German reflective items were provided, which were based on four context-appropriate items of the English likability scale of Reysen (2005): the virtual assistant is friendly/likable/warm/approachable (the complete list of German items as well as the results of the confirmatory factor analysis described below are available on request from the authors). Items were measured using a seven-point Likert-type scale with answers ranging from "strongly disagree" to "strongly agree". As a manipulation check, the respondents were asked to indicate the virtual assistant's perceived gender using a seven-point semantic differential scale using female and male on the endpoints.

Data Collection

Participants were recruited at two German universities, including bachelor students of an introductory course of information systems as well as master students. Additionally, we recruited doctoral students in the business-administration-related fields at one of those universities. All participants were able to enter a raffle to win one of four 25 € Amazon gift certificates as an incentive. We obtained a total of 54 completed questionnaires, of which four had to be removed because of failed manipulation and attention checks as well as one because of technical issues. An overview of the participants' demographics and condition

[3] Google Cloud TTS Service, which is based on the Google Cloud Text-to-Speech API and converts text into natural human speech, was used to generate the recordings. More specifically, we used the default settings of the German "WaveNet language C" as female voice, and "WaveNet language D" with an adjusted pitch of -4.00 as male voice.

assignments can be found in Table 2. 25 subjects received answers from the male-voiced virtual assistant, 24 from the female-voiced assistant. There were no statistically significant differences between experimental conditions considering the age (t-test) or gender (Fisher's exact test) of the participants. We can therefore assume that gender and age di not influence the comparison between groups.

	VA.Male N=25	**VA.Female** N=24	**Complete Sample** N=49	**p**
Age				
Mean	23.96	23.92	23.94	.963[a]
Standard Deviation	3.18	3.35	3.23	
Gender				
Male	14	12	26	.778[b]
Female	11	12	23	

a = Result of a t-test.
b = Result of a Fisher's exact test.

Table 2. Demographics and Controls

Results

Validation and Descriptives

The scales used were validated by using confirmatory factor analysis. Two items had to be removed because of insufficient factor loadings and, thus, four items remained for analysis, all of them achieving factor loadings of above .72. Since the average variance extracted (AVE) was .82, we were able to assume convergent validity. Additionally, Cronbach's alpha of our resulting scale was 0.84, indicating a solid reliability. An overview of the item scores as well as of the resulting composite scores can be found in Table 3. As described above, all items were measured on a seven-point Likert-type scale. The female voiced virtual assistant scored higher on every single item average of the perceived likability scale and the composite score of the construct was .73 points higher for the female-voiced assistant (5.35 vs. 4.62). In total, while both virtual assistants were considered rather likable, the female-voiced assistant clearly seemed to outperform the male-voiced counterpart in that regard.

	V.Male			**VA.Female**			**Complete Sample**		
	M	SD	Mdn	M	SD	Mdn	M	SD	Mdn
Perceived Likability*	**4.62**	**.98**	**4.50**	**5.35**	**1.02**	**5.75**	**4.98**	**1.06**	**5.00**
Item 1	4.20	1.12	4.00	5.17	1.17	5.50	4.67	1.23	5.00
Item 2	5.04	.93	5.00	5.42	1.14	6.00	5.22	1.05	5.00
Item 3	4.60	1.08	5.00	5.42	0.97	6.00	5.00	1.10	5.00
Item 4	4.64	1.47	5.00	5.42	1.21	6.00	5.02	1.39	5.00

M = mean, SD = standard deviation, Mdn = median
*=composite score, normalized with item count (=4)

Table 3. Item and Construct Descriptives

Hypothesis Testing

Group differences were then tested by using the Mann-Whitney U test, since a normal distribution of data necessary for the t-test could not be assumed (e.g., Field 2009). The result of the test can be found in Table 4.

Construct	Comparison	U	Significance	Effect size r
Perceived Likability	VA.Male/VA.Female	173.5	.006	.36

Table 4. Mann-Whitney U Test

We found a highly significant difference in perceived likability between the male-voiced and female-voiced assistants with a medium effect size ($U = 173.5$, $p = .006$, $r = .36$). More precisely, users perceive the female-voiced personal assistant to be more likable than the male-voiced variant. We therefore find support for our initial hypothesis.

Conclusion

In this study, we examined the impact of gender stereotyping on the perceived likability of virtual assistants. We conducted a controlled experiment with 49 participants, manipulating the perceived gender of the assistant by using differently pitched voices. We found that female-voiced virtual assistants are perceived to be more likable than their male-voiced counterpart, even if no difference in their ability exists.

These findings contribute to the understanding of the influence of gender stereotyping on the perceptions of human-computer interactions. Furthermore, our study emphasizes that CASA is still valid for modern applications and may offer further important insights into the understanding of personal virtual assistants. Taken with the results by Ernst and Herm-Stapelberg (2020), this study shows that it is important to evaluate the interaction with personal virtual assistants both from a utilitarian perspective and a hedonic perspective.

These findings hold important practical implications: First, if a virtual assistant is not perceived as being likable, users may refrain from adopting the underlying device. Especially in the context of smart speakers, this has serious consequences since the integrated virtual assistants are usually the only way to use them at all. Since this is the case, the sales of smart speakers are strongly dependent on people's likability perception of the virtual assistants. Since we found that female-voiced virtual assistants are perceived to be more likable than male-voiced virtual assistants, using the default setting of a female-voice might prove beneficial for the companies doing so. Our findings also emphasize that companies need to be aware of gender stereotypes when designing and developing virtual assistants (cf. Ernst and Herm-Stapelberg 2020). Finally, our findings underline that it is necessary for companies offering virtual assistants to consider more than just the utilitarian abilities of the assistants: They must also consider their hedonic-based traits.

This study has some limitations: First, since we only used two different German voices that were generated by the Google Text-to-Speech API, we do not know how users may react to other voices such as Alexa's or Siri's. It is therefore necessary to repeat the study with a variety of different voices in different languages, in order to draw more general implications. Second, the participants of our study included only rather young people located in Germany, which may influence the applicability of our findings. Indeed, different results may be found if the study were conducted in different countries, cultures, and age groups.

Since we found additional support for CASA and the impact of gender stereotypes in the context of virtual assistants, it would be interesting for future research to further explore the implications of vocal cues on a variety of dimensions of user experience, such as the credibility of a virtual assistant. Similarly, the usage of technologies such as chatbots, which do not rely on voice but rather on visual cues, e. g., an image depicting a female or male person, may also be influenced by gender stereotypes and subsequent studies should thus take a dedicated look at them, too. Moreover, since virtual assistants are also commonly used in the context of smartphones rather than smart homes, we also plan to evaluate whether the results hold

under these conditions. Finally, future research should also focus on the differences of the impact of gender stereotyping on the interaction with virtual assistants in different countries, cultures, and age groups.

References

Best, D. L. 2004. "Gender Stereotypes," in *Encyclopedia of Sex and Gender: Men and Women in the World's Cultures,* C. R. Ember and M. Ember (eds.), New York, NY: Kluwer Academic, pp. 11-23.

Carpenter, J., Davis, J. M., Erwin-Stewart, N., Lee, T. R., Bransford, J. D., and Vye, N. 2009. "Gender Representation and Humanoid Robots Designed for Domestic Use," *International Journal of Social Robotics* (1), pp. 261-265.

Damen, N., and Toh, C. 2019. "Designing for Trust: Understanding the Role of Agent Gender and Location on User Perceptions of Trust in Home Automation," *Journal of Mechanical Design* (141:6).

Davis, F. D., Bagozzi, R. P., and Warshaw, P. R. 1992. "Extrinsic and Intrinsic Motivation to Use Computers in the Workplace," *Journal of Applied Social Psychology* (22:14), pp. 1111-1132.

Edwards, C., Edwards, A., Stoll, B., Lin, X., and Massey, N. 2019. "Evaluations of an Artificial Intelligence Instructor's Voice: Social Identity Theory in Human-Robot Interactions," *Computers in Human Behavior* (90), pp. 357-362.

Ernst, C.-P. H., and Herm-Stapelberg, N. 2020. "Gender Stereotyping's Influence on the Perceived Competence of Siri and Co.," *HICSS 2020 Proceedings*.

Eyssel, F., and Hegel, F. 2012. "(S)He's Got the Look: Gender Stereotyping of Robots," *Journal of Applied Social Psychology* (42:9), pp. 2213-2230.

Eyssel, F., Kuchenbrandt, D., and Bobinger, S. 2011. "Effects of Anticipated Human-Robot Interaction and Predictability of Robot Behavior on Perceptions of Anthropomorphism," *HRI 2011 Proceedings*, pp. 61-67.

Field, A. 2009. *Discovering Statistics Using SPSS*, 3rd ed. ed., Thousand Oaks, CA: Sage.

Gough, H. G., and Heilbrun, A. B. 1983. *Adjective Check List Manual*, 2nd ed. ed., Palo Alto, CA: Consulting Psychologists Press.

Nass, C., and Lee, K. M. 2000. "Does Computer-Generated Speech Manifest Personality? An Experimental Test of Similarity-Attraction," *CHI 2000 Proceedings*, pp. 329-336.

Nass, C., Moon, Y., and Green, N. 1997. "Are Machines Gender Neutral? Gender-Stereotypic Responses to Computers with Voices," *Journal of Applied Social Psychology* (27:10), pp. 864-876.

Nass, C., Steuer, J., and Tauber, E. R. 1994. "Computers Are Social Actors," *CHI 1994 Proceedings*, pp. 72-78.

Pulles, N. J., and Hartman, P. 2017. "Likeability and Its Effect on Outcomes of Interpersonal Interaction," *Industrial Marketing Management* (66), pp. 56-63.

Reysen, S. 2005. "Construction of a New Scale: The Reysen Likability Scale," *Social Behavior and Personality* (33:2), pp. 201-208.

Schneider, A. K., Tinsley, C. H., Cheldelin, C. H., and Amanatullah, E. T. 2010. "Likeability V. Competence: The Impossible Choice Faced by Female Politicians, Attenuated by Lawyers," *Duke Journal of Gender Law & Policy* (17:2), pp. 10-42.

Sherriffs, A. C., and McKee, J. P. 1957. "Qualitative Aspects of Beliefs About Men and Women," *Journal of Personality* (25:4), pp. 451-464.

Tay, B., Jung, Y., and 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), pp. 75-84.

Williams, J. E., and Best, D. L. 1990. *Measuring Sex Stereotypes: A Multination Study*, Thousand Oaks, CA: Sage.

Williams, J. E., Satterwhite, R. C., and Best, D. L. 1999. "Pancultural Gender Stereotypes Revisited: The Five Factor Model," *Sex Roles* (40:7/8), pp. 513-525.