December 2021

User Perceptions and Stereotypic Responses to Gender and Age of Voice Assistants

Heather Watkins
Clemson University, hmwatkins1@gmail.com

USER PERCEPTIONS AND STEREOTYPIC RESPONSES TO GENDER AND AGE OF VOICE ASSISTANTS

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements of the Degree
Master of Science
Human Factors Psychology

---

by
Heather Watkins
October 2021

Accepted by:
Dr. Richard Pak, Committee Chair
Dr. Patrick Rosopa
Dr. Kelly Caine

ABSTRACT

Technologies such as voiced automation can aid older adults aging in place by assisting with basic home and health tasks in daily routines. However, currently available voice assistants have a common design - they are vastly represented as young and female. Prior work has shown that humans apply stereotypes to human-computer interactions similarly to human-human interactions. When these stereotypes are activated, users may lose trust or confidence in the device or stop using it all together. The purpose of this study was to investigate if users can detect age and gender cues of voiced automation and to understand the extent to which gender, age, and reliability elicit stereotypic responses which were assessed using history-based trust. A series of health-related voice automation scenarios presented users with voice assistants varying in gender, age, and reliability. Results showed differences in age and gender perceptions across participant age groups but no differences for overall trust. A three-way interaction showed that when voiced automation reliability was low, participants rated the young female voice assistant as significantly more trustworthy than all other voice assistants. This work contributes to our understanding of how anthropomorphic characteristics like age and gender in emerging technologies can elicit varied trust responses from younger and older adults.

TABLE OF CONTENTS

Page

Table of Contents (Continued)

Page

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Aging in place is defined as the “ability to live in one’s own home and community safely, independently, and comfortably, regardless of age, income, or ability level" (CDC, 2013). Surveys show that nine out of ten older adults plan to remain in their homes as they age (AARP, 2012). Doing so increases one’s quality of life, enables one to improve their physical and mental health, and allows maintaining social relationships (Black, 2008). Technology can aid older adults with daily tasks through smart home automation such as smart speakers (Vollmer & Ory, 2017). Smart speakers are a new class of consumer technology that combines highly anthropomorphized artificially intelligent agents that communicate to users via voice (Hoy, 2018). These devices can benefit older adults by assisting them with setting medication reminders, listening to the news, placing phone calls, or playing music (Vollmer & Ory, 2017). However, ultimate adoption and usage of such highly anthropomorphized technology depends on how users perceive the capabilities of that technology as well as their trust in the device. Past research indicates that these factors are likely to be heavily influenced by user stereotypes that they inevitably automatically apply to it (Pak, McLaughlin & Bass, 2014; Pak et al, 2012).

Human Characteristics of Smart Speakers

Smart speakers are wireless, hands-free devices that allow users to communicate with voice assistants by receiving voice input and delivering voice output. The devices require minimal set-up and are always on, enabling users to ask the voiced automation for what they need at any time. One characteristic that is shared among many smart speakers

is they often default to a female voice. For example, the four top voice assistants are Amazon's Alexa, Microsoft's Cortana, Apple's Siri, and Google Assistant, which all use female gendered voices.

As the design for today's smart speakers is proprietary, the decision to gender these voice assistants as female is mostly unknown. However, it may be motivated by past market data that suggests that individuals prefer the voices of females (Dong et al., 2020). For example, consumer researchers for a popular car manufacturer polled users to find out what voice to use for the first installed navigation system. They found that users tend to rate female voices more favorably than male voices (Griggs, 2011). Regardless of the previous rationale, as technology becomes more anthropomorphic (i.e., embodies human-like characteristics such as gendered, aged voices), human users will begin to apply pre-existing stereotypes to these devices (Pak et al., 2012).

Stereotypes in Human-Computer Interactions

Stereotypes are cognitive schemas about personality characteristics that are applied, often unconsciously, to others based on their group membership (Hamilton, 1979). A common example of this is that women have a warmer disposition. While this is an example of positive stereotypes, negative stereotypes are far more common and can skew our social perceptions of people based on factors like their race or gender. Such assumptions can generalize negative associations with certain groups, harboring feelings of mistrust and causing social ostracism (Dovidio et al., 2016). Similarly, prescriptive stereotypes can moderate our appraisal of other's capabilities, by causing us to project desirable characteristics onto individuals simply because they belong to a particular

group (Major, Mendes, & Dovidio, 2013). Examples of this include the application of benevolent sexism, which patronizes women who do not conform to the social gender expectations of warmth and dependence (Fiske, 2017). Ageism is another example of a prescriptive stereotype, whereby elderly people are viewed as subordinate and less competent (Cuddy, Norton, & Fiske, 2005). These examples illustrate the power of stereotypes and how broad, group-based associations can inform our appraisal of people at an individual level.

Interestingly, early research by Nass, Moon, and Green found that it is not only human subjects that fall victim to stereotyping (1997). In this study, participants received computer-based tutoring on one of two subjects before being asked to evaluate the competency of their tutor. The topics were geared towards either a masculine subject (computer and technology) or a feminine subject (love and relationships), and the computer voice was manipulated to sound either male or female. The results showed that overall, participants perceived the female-voiced computer as a better teacher on the topic of romantic relationships and the male-voiced computer as a better teacher on the subject of computers and technology ($F(1, 16) = 11.14$, $p < .01$). While the results demonstrated the presence of preexisting, gender-based stereotypes, a post-study debriefing indicated that participants unanimously stated that there was no difference between the female and male computers and denied harboring stereotypes. These unconscious differences mean that people have little insight into when and how their personal prejudices may be affecting their judgment in everyday life.

Building on these findings from computer-based teaching agents, recent literature has explored the application of stereotypes onto anthropomorphic robots. New research has found evidence of even more nuanced forms of stereotyping including making judgments about how reliable robots are when helping a human with a task. In one study by Pak et al. (2020), participants were shown a video-based vignette depicting a robot collaborating with a human to execute a task. Video scenarios varied the reliability (high & low) and age (younger & older) of the robot, as well as whether the collaboration was completed successfully or not. Results indicated that when the young robot successfully completed a cognitive task, it was rated as more trustworthy, whereas an older robot that completed the same task was trusted less. These findings are consistent with pre-existing stereotypes that people have about older adults as warm but not competent (Cuddy & Fiske, 2002). To summarize, existing literature has indicated the robust application of human stereotypes to technology. However, previous studies have exclusively manipulated expectations for reliability and trustworthiness through visual cues of age and gender (Nass, Moon, and Green, 1997; Pak et al., 2020). There is a gap in the literature regarding the stereotype-eliciting potential of non-visual aspects of group membership (e.g. voice).

It is clear that vocal cues, including the speaker's gender, can elicit stereotypes that align with those cues (Tay, Jung, & Park, 2014; Cambre & Kulkarn, 2019). It seems plausible then, that other characteristics of the voice can elicit other stereotypes. Huff et al. (2020) examined whether manipulations in the perceived age of a computer-generated voice could be detected and how this influenced assumptions about the speakers.

Participants were presented computer-generated audio reviews of automobiles, depicting either an older or younger voice. After listening to the clips, users were asked to rate the voices in regard to their perceptions of competency, informativeness, and age. The study found that computer-generated voices that had faster, higher-pitched speech were perceived as younger, while those that had a slower, lower-pitched speech were perceived as older. These findings suggest that users are able to accurately perceive and distinguish between age-based differences in artificial voices. What is still unclear is whether these cues will elicit age-based stereotypes and inform human-computer interactions similarly to what has been seen with gender-based cues (Lee, Nass, & Brave, 2000).

Study Rationale

The purpose of this study was to address the gap in existing literature regarding the application of age and gender-based stereotypes to human interactions with technology. Expanding on previous research that had focused on visual cues for gender and age (Pak, McLaughlin, & Bass, 2014; Pak et al., 2020), this study investigated the extent to which manipulation of vocal cues for age and gender can elicit stereotypic expectations during interactions with voiced automation. This study sought to build on the findings of Nass, Moon, & Green's (1997) and Huff et al. (2020) by simultaneously examining the effects of age and gender in this domain and evaluating their real-world application. The study's specific aims, objectives, and hypotheses are outlined below.

Aims and Hypotheses

The primary objectives of this study were: 1) to establish if both younger (age 18-23) and older (age 65-85) users can accurately detect the age and gender of a computer-

generated voice assistant 2) to identify the effect of user age on the perceived trust of the computer-generated voice assistants and 3) to examine the interaction of system reliability, perceived age of the voice, and perceived gender of the voice interact on overall perceptions of the voiced automation. Accordingly, this study's hypotheses were generated in line with prior research. The researchers hypothesized that 1) both older and younger adults would correctly report the age and gender of computer-generated voices (Huff et al., 2020), 2) Older users would report significantly higher overall levels of trust in comparison with younger users (Pak et al., 2014) and 3) When primed with expectations of low reliability, users will perceive younger female voice assistants as significantly less trustworthy than all other voice assistants (Pak et al., 2014).

METHOD

Study Design

The current experiment implemented a 2 (age of participant: younger/older) x 2 (age of voice assistant: younger/older) x 2 (gender of voice assistant: female/male) x 2 (automation reliability: low/high) mixed-factorial design. Participant age group was the quasi-independent grouping variable, within-group manipulations included voice assistant age, gender, and reliability, and trust was assessed as the dependent variable. A total of eight unique scenarios featuring different variations of age, gender, and reliability were randomly presented to participants, with each scenario being presented once (See Table 1).

Participants

A power analysis using the G*power computer program (Buchner, Faul, & Erdfelder, 1998) indicated that a total sample of 72 participants were needed to detect large effects (η2=.14) with 95% power using a mixed factor ANOVA. To account for attrition and possible uneven group sizes, data from 112 participants were recruited for the study. A total of 61 younger adults (44 females, M_{age}=19.16, *SD*=.92) were recruited from the university subject pool and received partial credit for a course requirement for their participation. An additional 51 older adults (28 females; M_{age}=71.18, *SD*=4.3) were recruited from the broader community and received monetary compensation for their participation.

Experimental Task

Researchers adapted the experimental task (see Figure 1) from previous studies (Pak et al., 2012, Pak et al., 2014), and modeled on typical user interaction with voice automation systems: asking for information. Online, participants were directed to ask simulated voice automation a set of specific questions, such as "What can I eat to increase my blood sugar levels", which was answered with the appropriate response. In each trial, a screen displaying an image of a smart speaker and a different question prompt was presented to the participant (full list of questions shown in Appendix A). In addition, the screen also displayed the past reliability of the voice assistant for that scenario (manipulated to be either 95% or 45% reliability). Participants were instructed to read the question on the screen aloud and then press "PLAY" when they were ready to hear a response. Once the participant selected "PLAY", an audio clip played the answer.

The voiced automation was manipulated to be either younger/older and either male/female.

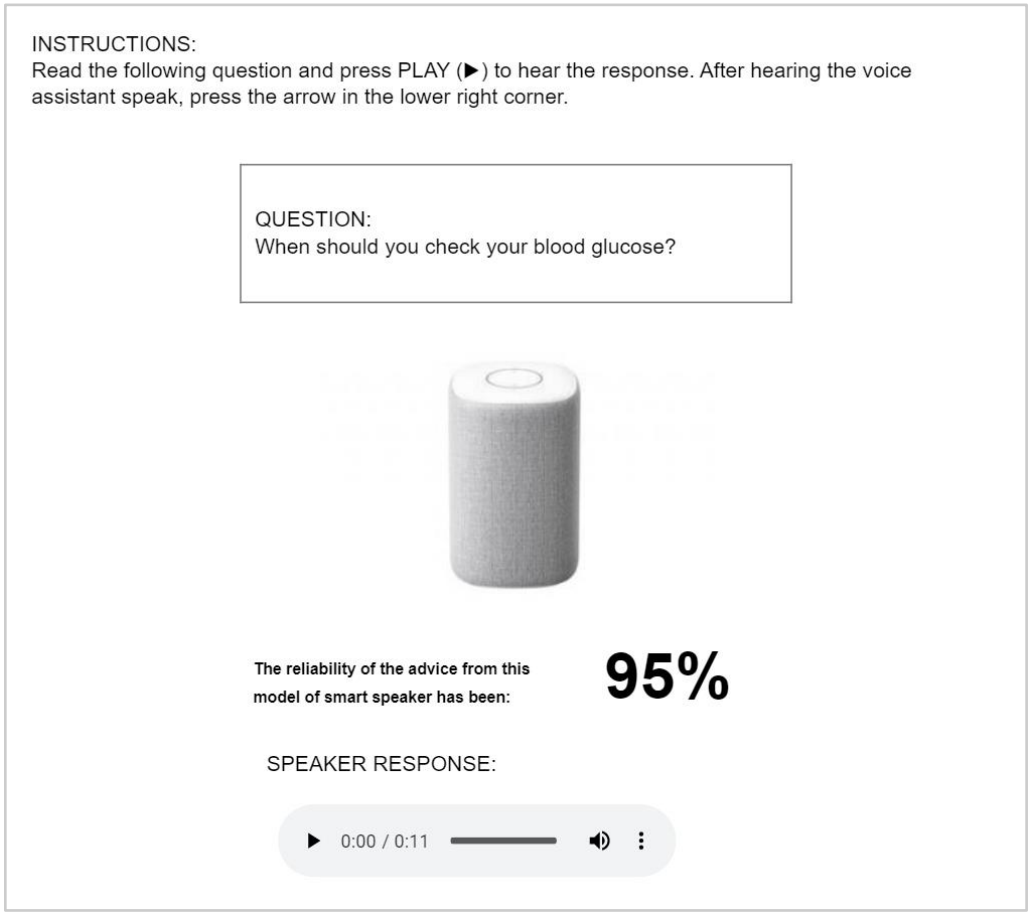

Figure 1. Experimental Task

Audio Stimuli

Original computer-generated voices portraying variations in gender (male, female), and age (younger, older) were created for this study. To achieve this, researchers used programs including language R, googleLanguageR, magrittr, dplyr packages, and Google's Clouds Text-to-Speech platform to varied pitch and speed in each variation of the voiced automation. For the older male voice, en-US-Wavenet-B was manipulated at a

rate of .60 and a pitch of -8. For the older female voice, en-US-Wavenet-E was manipulated at a rate of .48 and a pitch of -4.50. For the younger female voice, en-US-Standard-F was manipulated at a rate of .97 and a pitch of 3.5. Finally, the younger male voice, en-US-standard-I was manipulated at a rate of 1.07 and pitch of 4.50. The relevant R code used to generate audio clips has been included in Appendix B. Pilot data indicated that the voice samples were perceived in the expected age directions: younger male (*M*=18.1, *SD*=6.9), younger female (*M*=26.4, *SD*=5.2), older male (*M*=57.2, *SD*=11.1), and older female (*M*=52.5, *SD*=14.5). No gender estimation differences were found in pilot testing.

The usability testing scenario and health-related questions were adapted from a diabetes task used in a prior study (Pak et al., 2012). To manipulate reliability, each scenario conveyed past reliability of the speaker as either 45% or 95%, low and high respectively. The selected reliability percentages were informed by prior research detailing critical threshold points of reliability-induced automation complacency (Wickens & Dixon, 2007). The reliability of automation was only manipulated by informing participants of past reliability for the voiced automation portrayed in each vignette. The actual reliability of automation was held consistently at 100% across all vignettes. After each voice assistant interaction, participants answered a series of questions about their attitudes and perceptions.

Table 1. Voice Assistant Scenarios

Scenario	Reliability	Age (Voice Assistant)	Gender (Voice Assistant)
1	95%	Younger	Female
2	95%	Older	Female
3	95%	Younger	Male
4	95%	Older	Male
5	45%	Younger	Female
6	45%	Older	Female
7	45%	Younger	Male
8	45%	Older	Male

Study Measures

Researchers measured a total of five factors using a series of Likert-type scales as well as qualitative feedback. Measures were taken for the following: prior experience with technology, voice assistant frequency of use and usage type, trust, perceived age and gender, and reasoning criteria for trust ratings. Each measure is outlined below (See Appendix A for study questionnaires).

Prior Experience With Technology

The short-form Computer Proficiency Questionnaire, CPQ-12, was used to assess prior experience with technology (Boot et al., 2015). Participants rated 12 items using a 5-point Likert-type scale (1 = "*Never tried"*; 5 = "*Very easily*"). Example items include "I can use a computer keyboard to type" and "I can find information about my hobbies and interests on the Internet". Previous studies reported a reliability of Cronbach's α=0.95, while our study demonstrated a reliability of α=0.61.

Voice Assistant Use Frequency and Purpose

Participants were asked to report their frequency of use using a question adapted from the media and technology usage scale MTUAS (Rosen et al., 2013). Usage was reported on a 10-point scale (1="*Never*"; 10="*All the time*"). Participants were also asked to select typical reasons for use from a list including: "listening to music", "asking questions to obtain information", and "managing calendar".

Trust

As the primary measure of stereotype activation and use, two types of trust were assessed in this study: dispositional and history-based trust. Dispositional trust (i.e., an individual's likelihood to trust) was assessed before the study using the Automation Induced Complacency Potential Revised scale (AICP-R) (Merritt et al., 2019). Participants rated 10 items using a 5-point Likert-type scale (1 = "*strongly disagree"*; 5 = "*strongly agree*"), for scenarios including: "If life were busy, I would let an automated system handle some tasks" and "Constantly monitoring an automated system's performance is a waste of time". Previous studies reported a reliability range of Cronbach's α=0.79–0.87, while our study demonstrated a reliability of α = 0.66. History-based trust (i.e., trust after exposure to a system that is expected to show trust differences as a function of stereotype activation) was assessed after each trial using two questions adapted from Lee and Moray (1994). Participants indicated the degree to which they agreed with these statements using a 0 to 100 scale, where higher scores indicate higher levels of trust. The two questions were: "To what extent do you trust (i.e. believe in the accuracy of) the voice assistant in this scenario?" and "To what extent would you be

likely to follow the voice assistant's recommendation in this scenario?".

Perceived Age and Gender

Participants reported the perceived age and gender of each voice assistant through a manipulation check block that proceeded the task block of the experiment. Participants selected gender from choices of male and female, whilst age was indicated using a 0-100 scale.

Reasoning Criteria for Trust Ratings

After providing trust ratings for each voice assistant, participants were asked to provide a qualitative explanation for their reasoning for assigning the rating. Qualitative responses were grouped and analyzed using a grounded theory approach to establishing themes relating to criteria for trust ratings.

Procedure

All participants gave informed consent and completed demographic information prior to beginning the experiment. Initial measures were collected for technology experience, voice assistant usage, and trust in automation as described above. Next, experimental instructions for the task were provided and participants were informed that they were helping to test the usability of a health-focused smart speaker. Before the experiment began, all participants completed a practice trial to ensure that they were familiar with the task and understand the instructions. Following the practice trial, participants were informed that the experiment would now begin and were randomly presented with eight trials, each with a varied voice assistant scenario (see Table 1). After each trial, participants reported measures for perceived trust and provided information

regarding their selected trust rating. After data was collected for all possible scenarios, participants were asked to report the age and gender that they believed each voice assistant to be presenting. See Figure 2 for experimental procedures diagram.

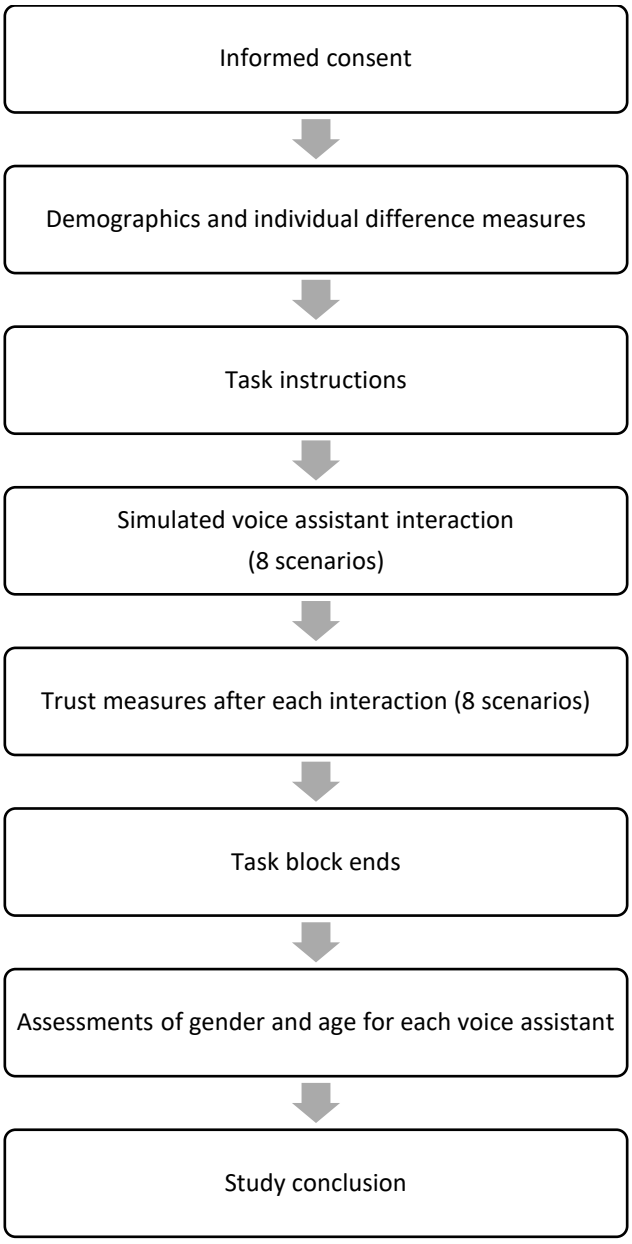

Figure 2. Experimental procedures diagram.

RESULTS

Data were collected for a total of 112 participants (younger = 61, older = 51) and included measures for the perceived trust of voiced automation across different age, gender, and reliability scenarios. The Mahalanobis distance metric (Tabachnick et al., 2007) was used for multivariate outlier detection and a single outlier was detected. Researchers made the decision not to exclude it from data analysis as there was no detectable change in the statistical significance of results when it was included in the analysis. Normality checks were conducted on all variables prior to statistical testing and showed multiple violations of the normality assumption. Given these initial findings, conservative estimates were used when interpreting the results of the following analyses.

Prior Experience With Technology

The pre-experiment Computer Proficiency Questionnaire provided insight into user competencies with using technology (Boot et al., 2015). Results were measured on a scale of 1 to 5, with higher scores indicating better competency. An independent samples t-test comparing younger and older participant CPQ scores showed no significant differences in younger (M=4.74, SD=.27) and older (M=4.63, SD=.54) adults' technology competency scores, $t(70.09)=1.39$, $p>.05$, $d=.28$. This finding shows that younger and older adult participants had similar levels of competency with using technology. Compared to the younger adult's mean competency, which is representative of their age group, the older adults' mean is less expected and may be a result of the recruitment process. Since advertisements for this study listed it as an online-based survey, we may have attracted older adults that are more technology-savvy than average.

Voice Assistant Use Frequency and Purpose

Voice assistant use frequency was measured on a 10-point scale indicating the frequency in which they use voice assistants, where higher values equal more frequent voice assistant usage. An independent samples t-test revealed significant differences in younger (M=4.31, SD=2.26) and older adults' (M=3.14, SD=2.49) frequency of voice assistant use, $t(110)= 2.62$, $p<0.05$, d=.49. These findings are unsurprising and are in line with available literature that found older adults tend to report having less perceived practical needs for automated technology, as well as finding it difficult to use them effectively (Trajkova & Martin-Hammond, 2020). Despite this distinction, when asked to select reasons for using a voice assistant from a list of options such as "checking the weather" or "playing music" (Hoy, 2018), an independent samples t-test found that there was no significant difference in the range of reasons for voice assistant use across participant age groups. This means that young (M=3.28, SD=2.48) and older adults' (M=4.47, SD=7.52), identify a similar range of usage, despite the later group reporting less frequent use $t(59.13)=-1.08$, $p>0.05$, d=.21.

Manipulation Check: Voice Assistant Age and Gender

At the end of each experimental block, participants were asked to report their estimations of age and gender for each voice assistant combination. An independent samples t-test was conducted to better understand individual differences in perceptions of voice assistant age. Significant differences were found in younger (M=17.08, SD=5.12) and older (M=26.24, SD=8.14) adults' perceptions of age in the younger male voice assistant conditions, $t(81.18)=-6.96$, $p<.001$, d=1.35. Specifically, younger adults

perceived the voice assistant to be significantly younger in age than older adult participants. Similarly, significant differences were also found in younger (M=46.31, SD=13.22) and older (M=52.14, SD=12.78) adults' perceptions of age in the older male voice assistant condition, t(107.68)=-2.37, p<.05, d=.45 with younger adults again perceived the voice assistant to be significantly younger in age than the older adult. Despite this, age estimations still fell in the respective voice assistant age categories of "younger" and "older" across their respective trials. Conversely, no significant differences were found in estimates of age for the younger or older female voice assistant conditions (all p values >.05). These findings suggest that both younger and older adults are better at distinguishing age in female voice assistants. This might be influenced by our current exposure to different female voiced automation including Amazon's Alexa, Microsoft's Cortana, Apple's Siri, and Google Assistant, which all use female gendered voices.

Chi-square tests were conducted to better understand age differences in perceptions of voice assistant gender. The results of the Chi-Squared test revealed that that older adults perceived the gender of young male voice assistants differently than younger adults, but these findings were not consistent across other trials, indicating that older adults perceive the gender of the younger male voice assistant differently than younger participants. No significant variances in estimations of gender were found for the other voiced automation conditions, p>.05.

Dispositional Trust

The pre-experiment survey measured participants' automation complacency using the AICP-R scale (Merritt et al., 2019). As a measure for an individual's propensity to trust automation, this scale was used here to establish dispositional trust. An independent samples t-test was conducted in order to determine age differences in dispositional trust. No significant differences were found between younger (M=3.31, SD=.52) and older adults' (M=3.39, SD=.52) dispositional trust ratings, $t(110)=-.823$, $p>.05$, d=.15. These findings suggest that no difference in younger and older adults' pre-existing perceptions of trust towards voiced automation. This is consistent with other findings showing no difference in younger and older adults' reported scores for automation complacency (Pak et al., 2020). However, it is worth noting that the AICP survey has not been approved for use with older adult subjects and therefore may have created some range issues.

History-Based Trust

Participants were asked to report their perceptions of trust after each voice assistant interaction. The analysis involved a 2(assistant reliability: low, high) x 2(assistant gender: male, female) x 2(assistant age: younger, older) x 2(participant age group: younger, older) mixed repeated-measures ANOVA. Since Box's equality test revealed a violation of equal variance assumptions, results were reported according to Pillai's trace as this test statistic has been shown to be more robust for errors in normality.

Significant main effects for history-based trust as a function of voice assistant reliability and voice assistant age were found. Specifically, lower overall history-based

trust scores were reported in low reliability trials (*M*=53.81, *SD*=18.83) compared to high reliability trials (*M*=75.62, *SD*=18.09) and (*F*(1,110)=142.11, *p*=<.001, ηp2=.564). With regard to voice assistant age, younger voice assistants received higher overall history-based trust ratings (*M*=71.09, *SD*=16.31) than older voice assistants (*M*=58.33, *SD*=19.6) and (*F*(1,110)=59.93, *p*< .001, ηp2=.353). Moreover, these effects were consistent across both participant age groups, indicating that participants from both age groups rated younger voice assistants to be more trustworthy than older voice assistants, and less reliable voice assistants as less trustworthy.

Gender of Voice Assistant

All participants reported estimations of voice assistant gender following experimental blocks. Statistical analysis showed no significant main effect for trust ratings as a function of voice assistant perceived gender were found (all *p* values >.05). In addition, no significant main effect for trust as a function of participant age group was found (*p*-value >.05). These findings contradict the predictions for hypothesis 2 that older adult participants would exhibit overall higher history-based trust towards voice assistants. However, a significant 2-way interaction was found for reported trust as a function of voice assistant reliability and gender, (*F*(1,110)=5.5, *p*=<.05, ηp2=.048). Follow-up pairwise comparisons revealed significant differences in history-based trust ratings as a function of reliability across both male and female voice assistant trials, with high reliability males (*M*=76.99, *SD*=17.83), and high reliability females (*M*=74.25, *SD*=21.99) being rated as significantly more trustworthy than low reliability male (*M*=52.72, *SD*=22.39) and low reliability females (*M*=54.9, *SD*=20.46).

Another significant 2-way interaction was found for gender and age of voice assistant ($F(1,110)=14.07$, $p< .001$, $\eta p2=.113$). Specifically, pairwise comparisons revealed significant differences in history-based trust ratings for all voice assistant gender and age combinations: young male voice assistants ($M=68.97$, $SD=17.1$), young female ($M=73.22$, $SD=20.1$), older male ($M=60.74$, $SD=21.97$), and older female ($M=55.93$, $SD=21.97$). These findings indicate that participants reported significantly different levels of trust for all gender and age combinations of voice assistants. Significant 3-way interactions included trust rating as a function of voice assistant reliability, voice assistant gender, voice assistant age, and participant age group. These interactions are presented in the graphs below.

Voice Assistant Reliability, Voice Assistant Age, and Voice Assistant Gender

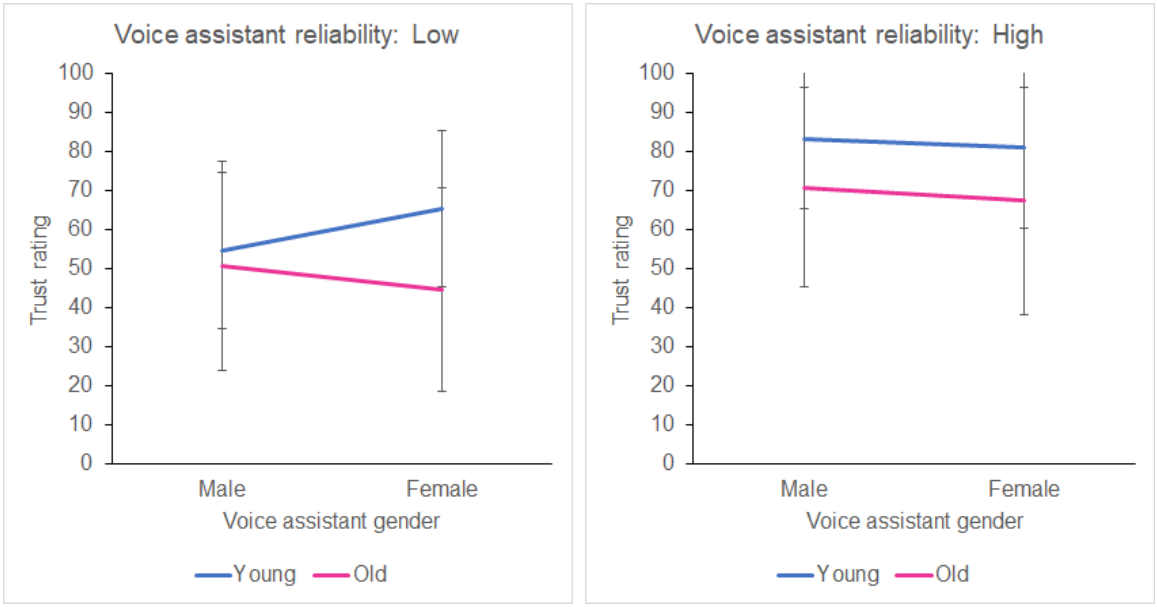

Figure 3. Trust ratings as a function of voice assistant age and gender in low and high reliability trials.

Figure 3 shows trust ratings as a function of voice assistant age and gender across low and high reliability trials. Data analysis showed a significant 3-way interaction was revealed between voice assistant gender, voice assistant age, and voice assistant reliability, $F(1,110)=8.2$, $p< .005$, $\eta p2=.070$). Specifically, follow up pairwise comparisons revealed that when voiced automation reliability was low, the younger female (M=65.25, SD=27.01) voice assistant received the highest history-based trust ratings, with significantly lower history-based trust ratings reported for than younger male (M=54.74, SD=28.11), older female (M=44.54, SD=26.1), and the older male (M=50.71, SD=26.77) voice assistants when reliability was low. For hypothesis 3, researchers anticipated a three-way interaction of system reliability, assistant gender, and assistant age. The researchers expected that when system reliability was low, users would perceive the younger female voice assistant as significantly less trustworthy than all other voice assistants. Contrary to expected findings for hypothesis 3 and prior findings in the literature (Pak et al., 2014), when system reliability was low, the younger female voice assistant received the highest history-based trust ratings. This finding shows that when voiced automation is unreliable, users are more likely to rely on a young female voice.

Given the market trend of a default young female voice for smart speakers and this empirically demonstrated greater trust in young female voice assistants, we conducted an exploratory analysis to better understand the possibility of exposure as a covariate of trust. Our measure of frequency of voice assistant usage was selected as the

covariate indicator of exposure. This analysis revealed no significant findings in our data for frequency of voice assistant usage as a covariate for trust in voice assistants, p>.05.

Participant Age, Voice Assistant Gender, and Voice Assistant Reliability

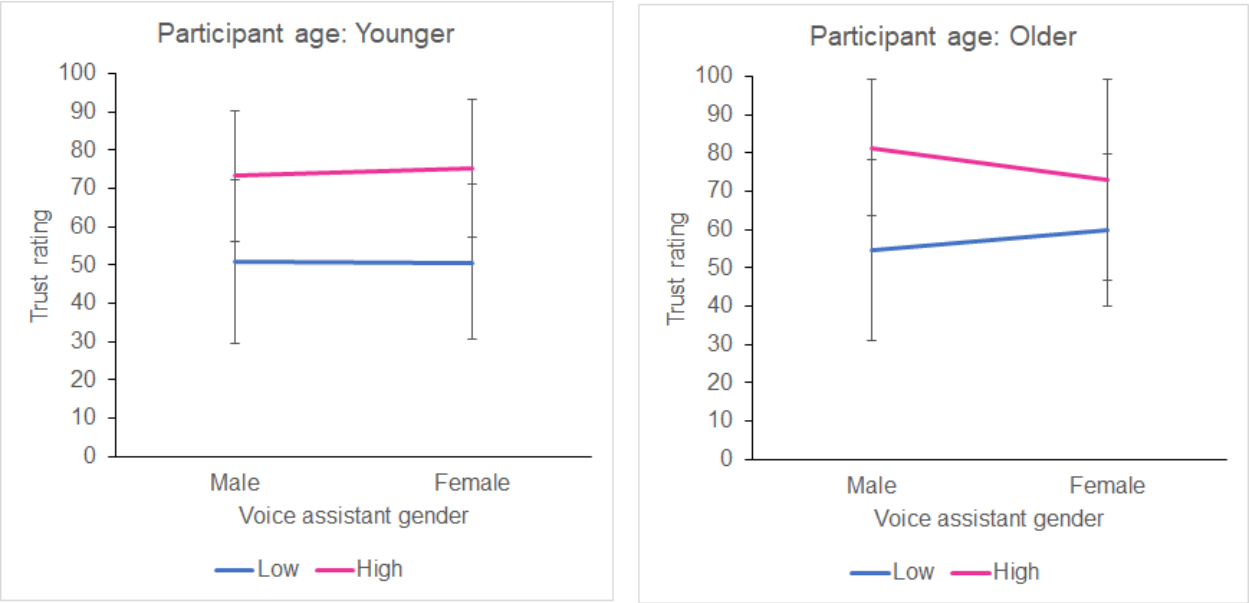

Figure 4. Trust ratings as a function of voice assistant gender and reliability in younger and older participant groups.

Figure 4 shows trust ratings as a function of voice assistant gender and reliability for younger and older participant groups. Data analysis showed significant 3-way interaction between participant age, voice assistant gender, and voice assistant reliability ($F(1,110)=10.73$, $p<.001$, $\eta p2=.089$). Pairwise comparisons revealed significant differences in history-based trust rating as a function of reliability between the two participant age groups. Specifically, the finding shows that when system reliability was low, older adults trusted the female voice assistant significantly more ($M=59.75$, $SD=19.87$) than younger adults with the same female assistant ($M=50.84$, $SD=20.20$).

Moreover, a significant difference in history-based trust rating was also observed for high reliability trials using a male voice assistant, in which older adults (*M*=81.48, *SD*=17.78) trust male voice assistants more than younger adults (*M*=73.23, *SD*=17.13) when reliability was high.

Participant Age, Voice Assistant Age, and Voice Assistant Reliability

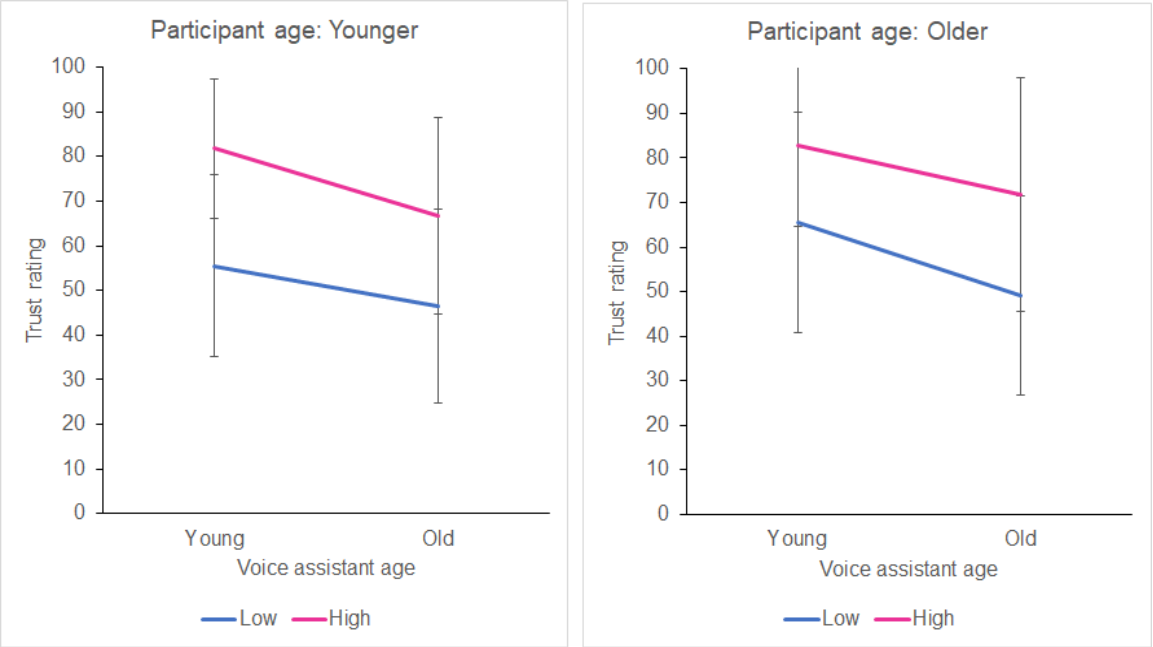

Figure 5. Trust ratings as a function of voice assistant age and reliability in younger and older participant groups.

Figure 5 indicates trust ratings as a function of voice assistant age and reliability for younger and older participant groups. Analysis revealed a significant 3-way interaction between participant age, voice assistant age, and voice assistant reliability ($F(1,110)=4.6$, $p<.001$, $\eta p2 = .04$). Specifically, young participants reported lower history-based trust (*M*=55.48, *SD*=20.43) compared to older adults' (*M*=65.4, *SD*=24.73)

when reliability was low for younger voice assistants, where older adults reported higher history-based trust. This result shows that when system reliability was low, older adult participants were significantly more trusting of younger voice assistants than younger adult participants were of the same younger assistants.

Qualitative Justification for Trustworthiness Rating

After providing a rating for history-based trust, participants were asked to elaborate on their reasoning for their reported ratings. Qualitative responses were analyzed using a ground theory approach and coded according to the nature of the criteria used to judge the voice assistants when selecting a trust rating. Four main categories of reasoning criteria were identified: dispositional trust, perceived confidence, personal knowledge, and explicitly stated reliability (See Table 2).

Table 2 Comparison of Reasoning Criteria for Trust By Age Group

Trust Criteria	Younger adult participants (*n*=61)	Older adult participants (*n*=51)
Dispositional trust	1	2
Perceived confidence	25	17
Personal knowledge	17	25
Explicitly stated reliability	18	7

Dispositional trust was categorized as any statements of participants' pre-existing attitudes toward automation, for example: "I never fully trust these devices so I would go and look it up myself", and "I don't usually trust what they say, so if I really need to

know I always go back and try to figure it out myself". Perceived confidence was denoted by the specific mentioning of how confident the voiced automation sounded. Example responses included: "Speaker did not sound sincere and confident" and "The lady sounds very smart and she sounds like the voice actors for medicine commercials which makes me trust her response more in this situation". Personal knowledge also played a role in assessing trustworthiness, with participants integrating their own knowledge in order to verify information from the voiced automation. When asked to elaborate, one participant said: "I already know that's the correct thing to do." The fourth category for justifying trustworthiness was through explicitly stating reliability, which was reportedly based on the presented percentage reliability during the voice assistant scenario. Example responses included: "The reliability is 95% so it is trustworthy.", "The reliability is only 45%", and "I wouldn't trust that automation to answer a question for me".

The three-way interaction of system reliability, voice assistant age, and voice assistant gender that was identified during data analysis was of particular interest to researchers. All participants reported significantly higher trustworthy scores for the young female voice assistant than all other voice assistants when system reliability was low. The qualitative responses indicated that participants assessed the young female voice assistant "I think that this voice assistant is very sure in what was stated", "Sounds more confident just needs to be more assertive", "She seemed confident in her answer", and "The upbeat voice sounds sure and similar to those in medicine commercials which makes me feel like in this scenario that the response is spot on".

DISCUSSION

The recent increase in human-automation interactions has clear applications for smart assistant integration for the aging-in-place population. A recent survey reported that 16.5 percent of the American population was 65 years of age or older and that percentage is expected to rise to 22 percent by 2050 (Statista Research Department, 2021). Their utility has already been established for assisting elderly individuals with routine tasks by setting medication reminders and helping place phone calls (Vollmer & Ory, 2017). As these smart-assistants become increasingly anthropomorphized, researchers have established that variations in observable factors like the gender and reliability of the smart speaker can elicit stereotypes that align with those cues (Tay, Jung, & Park, 2014; Cambre & Kulkarn, 2019; Park et al., 2020). Building on this research, the overall aim of this study was to examine how variations in vocal cues for the gender and age of voiced automation influenced the prevalence of stereotypic responses in younger and older users, and how these stereotypes affected users perceived reliability of the voice automation. Three research hypotheses were outlined for this study.

Hypothesis 1 predicted that both younger and older users would accurately perceive differences in age and gender across voice assistant trials. Results showed that younger users could accurately identify voice assistant age and gender across all trial conditions. In addition, older adults were able to accurately distinguish between younger and older females. However, older users provided significantly higher age estimations for younger and older male voice assistant trials, despite predictions still falling in the

appropriate direction of ‘younger’ and ‘older’. Moreover, young male voice assistant trials received the most variation in predictions for age and gender from older users.

These differences in findings may be due to age-related differences in hearing function caused by natural physiological changes that accompany the aging process. However, it is more likely that older users are better able to distinguish between older and younger, female voiced automation due to exposure to current voiced automation like Amazon’s Alexa, Microsoft’s Cortana, Apple’s Siri, and Google Assistant, which all use female-gendered voices. Regardless of rationale, the older voice assistants were still perceived in a lower age range (*M*=46.31 - 52.14) than would be considered “older” by most standards, which often report 65 as the low end of the age range for “older” adults. Future research should look closely to literature regarding age-related differences in voice acoustics with aims to inform vocal manipulations beyond pitch and speed that may influence perceptions of age in computer-generated voices.

Hypothesis 2 predicted in line with prior findings, that older adults would report significantly higher overall levels of trust than younger adults (Pak et al., 2014). Data analysis revealed no significant main effect of participant age group on trust rating in voice assistants. This null effect may be explained by the emerging beliefs of the nature of voice assistant usage and the companies that are behind their rollout. Newer technology may be met with skepticism from users across all ages. Other factors affecting perceptions of trust include user perceptions of privacy surrounding smart assistant usage. A recent poll showed that as many as 41% of voice assistant users have reservations about privacy, trust, and unauthorized listening from their devices (Olson & Kemery,

2019). Future iterations of similar research should consider implementing a pre-experiment measure of perceived privacy related to voice assistants in order to establish any potential interactions of significant interest and prior beliefs.

Hypothesis 3, predicted an interaction of system reliability, perceived age, and gender of the speaker would influence user perceptions of trust in the voice assistants. Through the reliability manipulation, findings provided support of this hypothesis, indicating differences in trust across all combinations of voice assistant reliability, age, and gender. And while it was expected in line with prior findings, that users would perceive younger female voice assistants as less trustworthy (Pak et al., 2014), results supported the contrary. Instead, participants were significantly more trusting of the young female voice assistant when the reliability of the system was low. The interaction of perceived reliability, gender, and age and the observed effect on user perceptions of trust in voice assistants support the expectation that individuals apply human-human stereotypes to human-computer interactions.

Limitations and Future Directions

These findings add to the existing literature on human-computer interactions, and the emerging trends showing that perceived indicators of human qualities like age and gender can influence perceptions of trust in voice assistants across users. Furthermore, user perceptions of system trustworthiness can influence their willingness to adopt or reject a new type of automation or to discontinue the use of automation. As such, future technological advancements should consider how user characteristics including age and

gender may interact with attributes of automation design, particularly when they portray human-like characteristics.

One limitation in this study was in the design of the experimental procedure. Since this study took place entirely online, the participant's interaction with the smart speaker was simulated whereby the participants "asked" the voice assistant questions and had to manually start the speaker's response. Due to differences in the audio quality of home speakers, perception of voices may have varied across users. The act of having to manually start the response might have also influenced their perceptions of the voiced automation. Furthermore, this design, which was adapted from previous research that featured questions and answers exclusively about diabetes. Prior diabetes knowledge was not assessed in this study and therefore, it is unknown how the consumer health domain or personal understanding of diabetes may have affected user perceptions of trust.

Future studies should consider pre-experimental knowledge checks, as well as investigate trust in voice assistants across various settings and domains (e.g. automobiles, home automation, and home healthcare). In addition, future research should incorporate additional measures to establish the presence of stereotypic ascription towards anthropomorphized systems. For example, a recent study by Tolmeijer et al. similarly explored how differences in vocal cues for pitch and gender influence this and trust formation. The researchers were able to identify these effects by asking participants to rate the automated system based on 24 traits that were stereotypically either male or female (2021). These stereotypes were still applied even when the voice was gender-ambiguous but had little effect on perceived trust levels.

Conclusion

Overall, the results of this study add to the growing body of literature surrounding anthropomorphic stereotyping and perceived trust during human-automation interaction. Specifically, these findings provide new insights demonstrating that, in addition to visual cues, vocal cues for characteristics such as age and gender can influence younger and older adults perceptions of trustworthiness when using voiced automation. A recent survey reported that 16.5 percent of the American population was 65 years of age or older and that percentage is expected to rise to 22 percent by 2050 (Statista Research Department, 2021). It is clear that as the aging population continues to grow, so too does the need for independent, in-home care options. As everyday use of automation continues to increase on both a personal and practical basis, it is pertinent for researchers to continue working towards improving our understanding of what factors influence successful human-automation interactions. Furthermore, within the context of technology in healthcare, the increased utilization by early individuals for everyday home tasks and health needs, should drive future research towards providing design guidelines that will simplify implementation and support a better quality of life for those aging in place.

APPENDIX A

COMPUTER PROFICIENCY QUESTIONNAIRE – 12 (short-form)

This questionnaire asks about your ability to perform a number of tasks with a computer. Please answer each question by placing an X in the box that is most appropriate. If you have not tried to perform a task or do not know what it is, please mark "NEVER TRIED", regardless of whether or not you think you may be able to perform the task.

Scale: 1 – Never tried, 2 – Not at all, 3 – Not very easily, 4 – Somewhat easily, 5 – Very easily

Computer Basics

I can:

Use a computer keyboard to type
Use a mouse

Printer

I can:

Load ink into the printer

Fix the printer when paper jams

Communication

I can:

Open emails

Send emails

Internet

I can:

Find information about local community resources on the Internet

Find information about my hobbies and interests on the Internet

Calendar

I can:

Use a computer to enter events and appointments into a calendar

Check the date and time of upcoming and prior appointments

Entertainment

I can:

Use a computer to watch movies and videos

Use a computer to listen to music

VOICE ASSISTANT USE FREQUENCY AND PURPOSE

A voice assistant is a type of software that is activated by voice. Voice assistants can answer questions and complete tasks. Some common examples of voice assistants are Amazon's Alexa, Google's Assistant, or Apple's Siri.

Please indicate how often you use a voice assistant.

Scale: 1 – Never, 2 – Once a year, 3 – Several times a month, 4 – Once a week, 5 – Several times a week, 6 – Once a day, 7 – Several times a day, 8 – Once an hour, 9 – Several times an hour, 10 – All the time

What do you use a voice assistant for? Please select all that apply. If you do not use a voice assistant, please mark "Not applicable / I do not use a voice assistant".

Options: Listening to music, Getting the news, Phone calls, Checking weather, Shopping,

Messaging, Playing games, Clock (alarm, timer, reminders), Exercise, Listening to audio book, Managing calendar, Managing other devices, Asking questions to obtain/learn facts or information, Managing shopping or to-do lists, Not applicable / I do not use a voice assistant

AUTOMATION INDUCED COMPLACENCY POTENTIAL-REVISED

The following questions are about automation. Automation describes the process in which devices are used to carry out tasks without human intervention. Some everyday examples of automation are automatic cruise control, GPS navigation, and robotic vacuum cleaners. Please read each statement carefully and select the one response that you feel most accurately describes your views and experiences. There are no right or wrong answers. Please answer honestly.

Scale: 1 – Strongly disagree, 2 – Somewhat disagree, 3 – Neither agree nor disagree, 4 – Somewhat agree, 5 – Strongly agree

1. When I have a lot to do, it makes sense to delegate a task to automation.
2. If life were busy, I would let an automated system handle some tasks for me.
3. Automation should be used to ease people's workload.
4. If automation is available to help me with something, it makes sense for me to pay more attention to my other tasks.

5. Even if an automated aid can help me with a task, I should pay attention to its performance.
6. Distractions and interruptions are less of a problem for me when I have an automated system to cover some of the work.
7. Constantly monitoring an automated system's performance is a waste of time.
8. Even when I have a lot to do, I am likely to watch automation carefully for errors.
9. It's not usually necessary to pay much attention to automation when it is running.
10. Carefully watching automation takes time away from more important or interesting things.

HISTORY-BASED TRUST QUESTIONS

Sliding scale: 0 (not at all) – 100 (Extremely)

1. To what extent do you trust (i.e. believe in the accuracy of) the voice assistant in this scenario?
2. To what extent would you be likely to follow the voice assistant's recommendation in this scenario?

APPENDIX B

```
#load packages
library(googleLanguageR)
library(magrittr)
library(googleAuthR)
library(dplyr)

#
#call google service account API key for authentication
gl_auth("C:/Users/[redacted]")
#
#older male low reliability
#specify text
gl_talk("Avocados are lowest in carbohydrates.",
        #specify which voice
        name = "en-US-Wavenet-B",
        #modify rate
        speakingRate = "0.60",
        #modify pitch
        pitch = "-8.00",
        #set file save name and call media player
        output = "OM_LR.wav") %>% gl_talk_player()

#
#older male high reliability
#specify text
gl_talk("You should always have fast acting carbohydrates to treat low
blood
        glucose readings",
        #specify which voice
        name = "en-US-Wavenet-B",
        #modify rate
        speakingRate = "0.60",
        #modify pitch
        pitch = "-8.00",
        #set file save name and call media player
        output = "OM_HR.wav") %>% gl_talk_player()

#
#older female high reliability
#specify text
gl_talk("You should check your blood glucose before eating, before bedt
ime,
        and if you feel high or low",
        #specify which voice
        name = "en-US-Wavenet-E",
        #modify rate
```

```
        speakingRate = "0.48",
        #modify pitch
        pitch = "-4.50",
        #set file save name and call media player
        output = "OF_HR.wav") %>% gl_talk_player()

#
#older female low reliability
#specify text
gl_talk("You should eat 15 grams of fast acting
        carbohydrate.",
        #specify which voice
        name = "en-US-Wavenet-E",
        #modify rate
        speakingRate = "0.48",
        #modify pitch
        pitch = "-4.50",
        #set file save name and call media player
        output = "OF_LR.wav") %>% gl_talk_player()

#
#young female high reliability
#specify text
gl_talk("Common areas for insulin injections include the abdomen, arms,
and
        thighs",
        #specify which voice
        name = "en-US-Standard-C",
        #modify rate
        speakingRate = "1.07",
        #modify pitch
        pitch = "05.00",
        #set file save name and call media player
        output = "YF_HR.wav") %>% gl_talk_player()

#
#young female low reliability
#specify text
gl_talk("The first thing you should do if you feel low is check your
        blood glucose.",
        #specify which voice
        name = "en-US-Standard-C",
        #modify rate
        speakingRate = "1.07",
        #modify pitch
        pitch = "05.00",
        #set file save name and call media player
        output = "YF_LR.wav") %>% gl_talk_player()
```

```
#
#young male high reliability
#specify text
gl_talk("A healthy way to cope during a stressful time is to talk with
a friend or family member.",
        #specify which voice
        name = "en-US-standard-i",
        #modify rate
        speakingRate = "0.95",
        #modify pitch
        pitch = "04.50",
        #set file save name and call media player
        output = "YM_HR.wav") %>% gl_talk_player()

#
#young male low reliability
#specify text
gl_talk("Your blood glucose may go low if you are taking too much
insulin.",
        #specify which voice
        name = "en-US-standard-i",
        #modify rate
        speakingRate = "0.95",
        #modify pitch
        pitch = "04.50",
        #set file save name and call media player
        output = "YM_LR.wav") %>% gl_talk_player()
```

REFERENCES

AARP. *The United States of Aging Survey* (2012). Retrieved January 28, 2020, from https://www.aarp.org/livable-communities/learn/research-trends/info-12-2012/the-united-states-of-aging-2012.html

ADI - State of Voice Assistants. (2018, September 10). Retrieved January 28, 2020, from https://www.slideshare.net/adobe/adi-state-of-voice-assistants-113779956

Black, K. (2008). Health and aging-in-place: Implications for community practice. *Journal of Community practice*, *16*(1), 79-95.

Boot, W. R., Charness, N., Czaja, S. J., Sharit, J., Rogers, W. A., Fisk, A. D., & Nair, S. (2015). Computer proficiency questionnaire: assessing low and high computer proficient seniors. *The Gerontologist*, 55(3), 404-411.

Buchner, A., Faul, F., & Erdfelder, E. (1998). A priori, post-hoc, and compromise power analyses for MS-DOS (German Version).

Cambre, J., & Kulkarni, C. (2019). One Voice Fits All? Social Implications and Research Challenges of Designing Voices for Smart Devices. *Proceedings of the ACM on Human-Computer Interaction, 3*(CSCW), 1-19.

Center for Disease Control and Prevention. (2013). Healthy places terminology: Aging in place. Retrieved February 15, 2020 from http://www.cdc.gov/healthyplaces/terminology.htm

Cerrato, L., Falcone, M., & Paoloni, A. (2000). Subjective age estimation of telephonic voices. *Speech Communication, 31*(2-3), 107-112.

Cuddy, A. J., & Fiske, S. T. (2002). Doddering but dear: Process, content, and function in stereotyping of older persons. *Ageism: Stereotyping and prejudice against older persons*, *3*(1), 26.

Dong, J., Lawson, E., Olsen, J., & Jeon, M. (2020, December). Female voice agents in fully autonomous vehicles are not only more likeable and comfortable, but also more competent. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 64, No. 1, pp. 1033-1037). Sage CA: Los Angeles, CA: SAGE Publications.

Dovidio, J. F., Eggly, S., Albrecht, T. L., Hagiwara, N., & Penner, L. A. (2016). Racial biases in medicine and healthcare disparities. *TPM: Testing, Psychometrics, Methodology in Applied Psychology*, *23*(4).

Edmondson, M. (2018). *googleLanguageR: Call Google's 'Natural Language' API, 'Cloud Translation' API, 'Cloud Speech' API and 'Cloud Text-to-Speech' API. R Package Version 0.2.0.* Available online at: https://CRAN.R-project.org/package=googleLanguageR (accessed October 21, 2020).

Griggs, B. (2011). Why computer voices are mostly female. CNN. (Retrieved February 09, 2020, from http://www.cnn.com/2011/10/21/tech/innovation/female-computer-voices/

Hamilton, D. L. (1979). A cognitive-attributional analysis of Stereotyping. In *Advances in experimental social psychology* (Vol. 12, pp. 53-84). Academic Press.

Hoy, M. B. (2018). Alexa, Siri, Cortana, and more: an introduction to voice assistants. *Medical reference services quarterly*, *37*(1), 81-88.

Huff, E., Stigall, B., Brinklet, J., Pak, R., & Caine, K. (2020) Can Computer-Generated Speech Have an Age? *Proceedings for CHI 2020*, in press.

Jian, J., Bisantz, A.M., Drury, C.G., & Llinas, J. (2000). Foundations for an Empirically Determined Scale of Trust in Automated Systems.

Jung, E. H., Waddell, T. F., & Sundar, S. S. (2016). Feminizing robots: User responses to gender cues on robot body and screen. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 3107-3113).

Kuchenbrandt, D., Häring, M., Eichberg, J., Eyssel, F., & André, E. (2014). Keep an eye on the task! How gender typicality of tasks influence human–robot interactions. *International Journal Of Social Robotics*, *6*(3), 417-427. doi:10.1007/s12369-014-0244-0

Lee, E. J., Nass, C., & Brave, S. (2000). Can computer-generated speech have gender? An experimental test of gender stereotype. In *Proceedings of the 2000 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 289-290).

Macrae, C. N., Stangor, C., & Hewstone, M. (Eds.). (1996). *Stereotypes and stereotyping*. Guilford Press.

Major, B., Mendes, W. B., & Dovidio, J. F. (2013). Intergroup relations and health disparities: A social psychological perspective. *Health Psychology*, *32*(5), 514.

Merritt SM, Ako-Brew A, Bryant WJ, Staley A, McKenna M, Leone A and Shirase L (2019) Automation-Induced Complacency Potential: Development and Validation of a New Scale. *Front. Psychol.* 10:225. doi: 10.3389/fpsyg.2019.00225

Merritt, S. M., & Ilgen, D. R. (2008). Not all trust is created equal: Dispositional and history-based trust in human-automation interactions. *Human Factors*, *50*(2), 194-210.

Mitchell, W. J., Ho, C. C., Patel, H., & MacDorman, K. F. (2011). Does social desirability bias favor humans? Explicit–implicit evaluations of synthesized speech support a new HCI model of impression management. *Computers in Human Behavior*, *27*(1), 402-412.

Nass, C., Moon, Y., & Green, N. (1997). Are machines gender neutral? Gender stereotypic responses to computers with voices. *Journal of Applied Social Psychology*, 27, 864-876.

Olson, C., & Kemery, K. (2019). Voice report: Consumer adoption of voice technology and digital assistants. *Microsoft: Seattle, DC, USA.*

Pak, R., Crumley-Branyon, J. J., de Visser, E. J., & Rovira, E. (2020). Factors that affect younger and older adults' causal attributions of robot behavior. *Ergonomics*, (just-accepted), 1-49.

Pak, R., Fink, N., Price, M., Bass, B., & Sturre, L. (2012). Decision support aids with anthropomorphic characteristics influence trust and performance in younger and older adults. *Ergonomics*, 55(9), 1059-1072.

Pak, R., McLaughlin, A. C., & Bass, B. (2014). A multi-level analysis of the effects of age and gender stereotypes on trust in anthropomorphic technology by younger and older adults. *Ergonomics*, *57*(9), 1277-1289.

Rosen, L. D., Whaling, K., Carrier, L. M., Cheever, N. A., & Rokkum, J. (2013). The media and technology usage and attitudes scale: An empirical investigation. *Computers in human behavior*, *29*(6), 2501-251

Ruffman, T.,Murray, J.,Halberstadt, J., & Vater, T. (2012). Age-related differences in deception. Psychology and Aging, 27, 543–549. doi: 10.1037/a00233801.

Statista Research Department. *U.S. - seniors as a percentage of the population 1950-2050*. Statista. (2021, January 20). Retrieved September 21, 2021, from https://www.statista.com/statistics/457822/share-of-old-age-population-in-the-total-us-population

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2007). *Using multivariate statistics* (Vol. 5, pp. 481-498). Boston, MA: Pearson.

Tay, B., Jung, Y., & Park, T. (2014). When stereotypes meet robots: the double-edge sword of robot gender and personality in human–robot interaction. *Computers in Human Behavior*, 38, 75-84.

Tolmeijer, S., Zierau, N., Janson, A., Wahdatehagh, J. S., Leimeister, J. M. M., & Bernstein, A. (2021, May). Female by Default?–Exploring the Effect of Voice Assistant Gender and Pitch on Trait and Trust Attribution. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1-7).

Trajkova, M., & Martin-Hammond, A. (2020, April). " Alexa is a Toy": Exploring Older Adults' Reasons for Using, Limiting, and Abandoning Echo. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1-13).

Tung, F. W. (2011, July). Influence of gender and age on the attitudes of children towards humanoid robots. In *International Conference on Human-Computer Interaction* (pp. 637-646).

Vollmer Dahlke, D., & Ory, M. G. (2017). Emerging opportunities and challenges in optimal aging with virtual personal assistants. *Public Policy & Aging Report*, *27*(2), 68-73.

Wickens, C.D., & Dixon, S.R. (2007). The benefits of imperfect diagnostic automation: A synthesis of the literature. *Theoretical Issues in Ergonomics Science, 8* (3), 201-212.

Xiang, H., Zhou, J., & Xie, B. (2020, July). Understanding Older Adults' Vulnerability and Reactions to Telecommunication Fraud: The Effects of Personality and Cognition. In *International Conference on Human-Computer Interaction* (pp. 351-363). Springer, Cham.