



# The Terms of “You(s)”: How the Term of Address Used by Conversational Agents Influences User Evaluations in French and German Linguaculture

Joseph Ollier<sup>1,2\*†</sup>, Marcia Nißen<sup>1,2†</sup> and Florian von Wangenheim<sup>1,2</sup>

<sup>1</sup> Chair of Technology Marketing, Department of Management, Economics and Technology (D-MTEC), ETH Zürich, Zurich, Switzerland, <sup>2</sup> Centre for Digital Health Interventions (CDHI), Department of Management, Economics and Technology (D-MTEC), ETH Zürich, Zurich, Switzerland

## OPEN ACCESS

### Edited by:

Shona D’Arcy,  
Independent Researcher,  
Dublin, Ireland

### Reviewed by:

Celine De Looze,  
Trinity College Dublin, Ireland  
Emre Sezgin,  
Nationwide Children’s Hospital,  
United States

### \*Correspondence:

Joseph Ollier  
jollier@ethz.ch

<sup>†</sup>These authors have contributed  
equally to this work and share first  
authorship

### Specialty section:

This article was submitted to  
Digital Public Health,  
a section of the journal  
Frontiers in Public Health

Received: 06 April 2021

Accepted: 03 December 2021

Published: 05 January 2022

### Citation:

Ollier J, Nißen M and von  
Wangenheim F (2022) The Terms of  
“You(s)”: How the Term of Address  
Used by Conversational Agents  
Influences User Evaluations in French  
and German Linguaculture.  
Front. Public Health 9:691595.  
doi: 10.3389/fpubh.2021.691595

**Background:** Conversational agents (CAs) are a novel approach to delivering digital health interventions. In human interactions, terms of address often change depending on the context or relationship between interlocutors. In many languages, this encompasses *T/V distinction*—formal and informal forms of the second-person pronoun “You”—that conveys different levels of familiarity. Yet, few research articles have examined whether CAs’ use of T/V distinction across language contexts affects users’ evaluations of digital health applications.

**Methods:** In an online experiment ( $N = 284$ ), we manipulated a public health CA prototype to use either informal or formal T/V distinction forms in French (“tu” vs. “vous”) and German (“du” vs. “Sie”) language settings. A MANCOVA and *post-hoc* tests were performed to examine the effects of the independent variables (i.e., T/V distinction and Language) and the moderating role of users’ demographic profile (i.e., Age and Gender) on eleven user evaluation variables. These were related to four themes: (i) Sociability, (ii) CA-User Collaboration, (iii) Service Evaluation, and (iv) Behavioral Intentions.

**Results:** Results showed a four-way interaction between T/V Distinction, Language, Age, and Gender, influencing user evaluations across all outcome themes. For French speakers, when the informal “T form” (“Tu”) was used, higher user evaluation scores were generated for younger women and older men (e.g., the CA felt more humanlike or individuals were more likely to recommend the CA), whereas when the formal “V form” (“Vous”) was used, higher user evaluation scores were generated for younger men and older women. For German speakers, when the informal T form (“Du”) was used, younger users’ evaluations were comparable regardless of Gender, however, as individuals’ Age increased, the use of “Du” resulted in lower user evaluation scores, with this effect more pronounced in men. When using the formal V form (“Sie”), user evaluation scores were relatively stable, regardless of Gender, and only increasing slightly with Age.

**Conclusions:** Results highlight how user CA evaluations vary based on the T/V distinction used and language setting, however, that even within a culturally homogenous language group, evaluations vary based on user demographics, thus highlighting the importance of personalizing CA language.

**Keywords:** conversational agents, chatbots, term of address, T/V distinction, linguaculture, digital health

## INTRODUCTION

### Designing Conversational Agents for Healthcare

Conversational agents (CAs) are intelligent computer programs that engage users in human-like conversations and include text-based chatbots, voice-activated assistants, and embodied conversational agents (1). The use of CAs in healthcare service delivery has become increasingly widespread as organizations and practitioners recognize their ability to transform the healthcare sector and empower individuals to co-manage their care effectively (2). A broad range of scientifically evaluated healthcare chatbots are currently (commercially) available, providing digital health solutions across the patient journey from diagnostic conversations of a regular doctor visit [e.g., BABYLON<sup>1</sup> and ADA<sup>2</sup> (3)], consultations on sensitive health topics (e.g., the HIV CHATBOT<sup>3</sup>), therapy for specific chronic diseases [e.g., mental health (4): WOEBOT<sup>4</sup> (5, 6), WYSA<sup>5</sup> (7), and TESS<sup>6</sup> (8–10); cardiovascular diseases: FLORENCE<sup>7</sup> (11)] to general lifestyle health [e.g., LARK<sup>8</sup> (12)]. Benefits of CA use in healthcare include improving availability, personalization, and efficacy of service delivery (13). Moreover, due to their highly scalable nature, CAs have been noted as a promising method to address health disparities between developed and developing nations and ensure equitable healthcare service delivery worldwide (14).

From initial investigations into the suitability of CAs to act as healthcare partners (15), research has now turned to understanding best design practices for the anthropomorphic user interfaces that CAs use (16, 17). Research has demonstrated how visual, conversational, and identity-related cues trigger “humanness heuristics” (18, 19) and affective states in users similar to natural human communication (20). Design factors such as physical appearance (21), gender (22, 23), and speech dialect (24) can be tailored to match users’ cultural and demographic background and help to establish rapport (15, 20) and perceptions of a CAs personality (25). Language-based cues are of particular importance due to their strong role in driving user engagement (1). For example, research has highlighted how CA use of task and social-based communication (15, 26, 27), politeness (28), interactivity (17, 18), and information quality (29) have been linked to user evaluation outcomes such as interpersonal closeness (22), intention to use (30), satisfaction (31, 32), trust (33), and user self-disclosure (34).

To date though, few research articles have examined important cultural- and sociolinguistic phenomena in CA design across diverse *linguacultures* (i.e., where language and culture constitute a single domain), and how these influence perceptions of CAs and their effectiveness in healthcare service delivery (14).

This is particularly important, however, as language has a strong impact on social cognition and the co-construction of meaning between dyadic conversational partners (35, 36). For example, in English-speaking contexts, *terms of address* such as “Sir” (37) or “Mate” (38) vary in contextual appropriateness depending on the focus of address (e.g., police officer, friend). In CA contexts using users’ first names as the term of address has been linked to increased perceptions of CA politeness and thoughtfulness, with the caveat that this may be bound to cultural limits and preferences (25). In scaling up digital-health interventions globally, it is therefore imperative to further investigate CA language phenomena such as the term of address in diverse language contexts (14, 39).

### Term of Address: Design Considerations

In the current study, we examine a particular term of address cue: *T/V distinction* or *Tu/Vous distinction* (40), which refers to the use of different second-person pronouns (“You”) in some languages, denoting a combination of less (T form) or more (V form) formality, distance, or emotional detachment (41). Arising originally from the Latin pronouns “*Tu*” and “*Vos*” (becoming “*Thou*” and “*You*” in English), T/V distinction is widely used in Indo-European languages (42), for example, “*du*” and “*Sie*” in German, “*tu*” and “*vous*” in French, “*tú*” and “*vosotros/vosotras/usted(es)*” in Spanish, with similar use in other non-Latin related languages such as Chinese, Malaysian, and Korean (43). T/V distinction is said to encode interactional meanings and shape normative expectations (44), such as politeness etiquette (41, 45, 46), which when breached by a communication partner may disrupt the cultural script in play (44), and be perceived as an insult (41), membership of a different social class (47), affiliation with another culture or grouping (48), and lead to outcomes such as customer dissatisfaction (43, 49, 50).

For designers of CAs, therefore, it remains vital to investigate CA T/V distinction usage to facilitate engaging user experiences (51). As individuals look for cues in the cultural script to orient themselves and understand potential outcomes, benefits, or goals of relationships (52), designers of CAs can provide clear and stable meaning by appropriate utilization of the T/V form for a given user group (45). In a wider public health context, ensuring the correct reception of CA-based technology can help extend healthcare service equitably across the world, and address some of the shortages of human resources for health and clinical services, to “improve accessibility, availability, affordability, and acceptability of public health services worldwide” (20). Yet, as highlighted by the World Health Organization, to do so, the research community must address the risk of design biases when developing new interventions for diverse cultural backgrounds (14) which when simply transferred from English-speaking contexts may apply a “cultural filtering” effect (43) and disregard important communicative nuances in cultural scripts of a given linguaculture (43, 49).

The current research, therefore, investigates CA use of T/V distinction in two unique linguacultures of French and German and explores the role T/V distinction plays in user evaluation of eleven outcome variables grouped into the following themes: (i)

<sup>1</sup><https://babylonhealth.com>

<sup>2</sup><https://ada.com/de/>

<sup>3</sup><https://bot.hiv.gov/>

<sup>4</sup><https://woebothealth.com/>

<sup>5</sup><https://wysa.io/>

<sup>6</sup><https://www.x2ai.com/>

<sup>7</sup><https://florence.chat/>; Cottrell et al. (11).

<sup>8</sup><https://www.lark.com/>; Stein and Brooks (12).

Sociability (i.e., Social Presence, and Conversational Enjoyment), (ii) CA-User Collaboration (i.e., CA Trust, Co-Production, Perceived Privacy Protection, and Privacy Concern), (iii) Service Evaluation (i.e., Perceived Ease of Use, Perceived Usefulness, and Service Satisfaction) and (iv) Behavioral Intentions (i.e., Intended Usage, and Net Promoter Score). Further information on these outcome variables is available in the **Supplementary Materials**.

## Hypotheses Development

Similarities and variations in the usage of T/V distinction between linguacultures have been demonstrated with regard to contextual appropriateness (45), the relationship between interlocutors (52), and subtleties in semantic meanings (43, 49). For French speakers, utilization of V form generally occurs more frequently in interactions as a method to convey a base level of respect for all including strangers (44, 53) as well as to exhibit respect to hierarchy (54). In German, V form is employed less frequently, with many forms of relationships progressing to T form immediately or after a short interaction (44). While T form occurs more frequently in German, its usage however typically offers less significance as a relationship marker, being more readily extended to acquaintances or strangers (44), and lacking as rich connotations of proximity, intimacy, and positive affect as in French-speaking settings (55). Nevertheless, contextual influences remain highly consequential (45), and in certain German-speaking settings, such as when dealing with customer complaints (56) or in interactions with police officers (53) T form usage can be deemed highly inappropriate to the point of provocation (57). Since the T form is more commonly used in German and the V form more commonly in French, we hypothesize that:

**H1a:** French speakers will exhibit a preference for a CA using the V form (i.e., “vous”)

**H1b:** German speakers will exhibit a preference for a CA using the T form (i.e., “du”).

In line with users’ stated (subjective) preferences, we further posit that T/V distinction will be found to (objectively) cause improved user evaluations (i.e., the CA will be rated more humanlike, or individuals will rate higher likelihood to recommend the CA) when utilizing the T/V distinction appropriate to the given linguaculture (German, French). Accordingly, we hypothesize that:

**H2a:** For French speakers, the use of the formal T/V distinction (i.e., V form) “vous” by CAs will improve individuals’ user evaluation scores.

**H2b:** For German speakers, the use of the informal T/V distinction (i.e., T form) “du” by CAs will improve individuals’ user evaluation scores.

While a linguaculture exhibits stable traits that are distinct from other linguacultures (44), within-linguaculture differences related to users’ age and gender also exist. Age has often been linked to T/V distinction appropriateness, as youth are typically more exposed to emergent cultural trends influencing linguistic evolution (44, 56). For example, older French individuals prefer V form (47), younger German speakers are more likely to use

T form (52), and older German speakers may even view T form usage without permission as provocative (56). Gender differences in T/V usage have also been exhibited with French-speaking men more likely to give T form (58) and receive V form than women in interactions with other interlocutors (47), which may correspond to wider differences in language use exhibited between men and women more generally (59) and shifting of gender roles through time (60).

Navigating evolutions in T/V distinction usage in French- and German-speaking linguacultures has often proved difficult for firms and organizations, with changes in T/V distinction used in the workplace and commercial settings (for example, imposition of T form) meeting with scrutiny or resistance (54–56). Additionally, as CAs operate in a channel that typically involves informal, bi-directional communication (i.e., instant text messaging) (61, 62), it is unclear to what extent general cultural norms for formality from strangers in professional settings (41) transfer to the digital environment, and whether variations can be evidenced based on demographic profiling. Thus, while T form has become more widely accepted (63), there remains a body of evidence demonstrating T/V distinction preferences are complex and multifaceted. Therefore, we also examine whether:

**H3:** User Age and Gender jointly moderate the relationship between T/V distinction and Language and user evaluation scores.

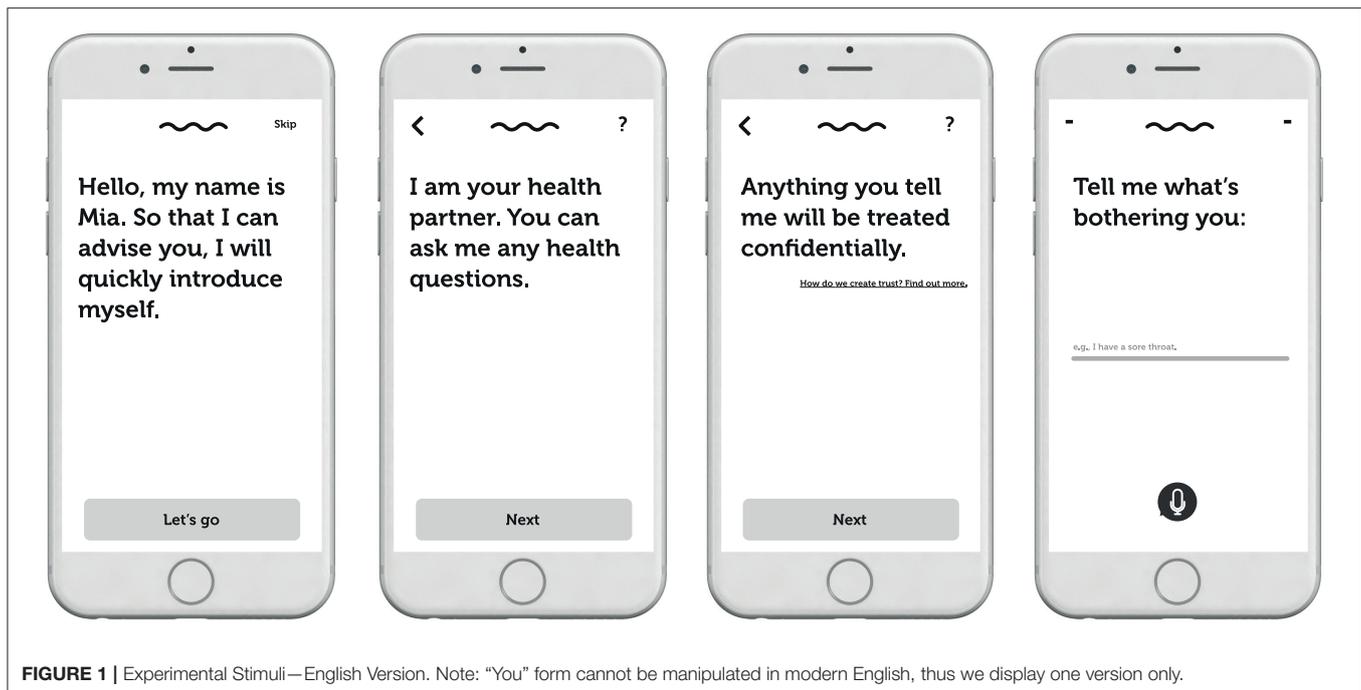
## MATERIALS AND METHODS

### Study Design

To investigate our hypotheses, we conducted a web-based experiment that examined users’ preferences for either the T or V form (H1a and H1b), the effects of a healthcare CA’s T/V distinction use in two language contexts (H2a and H2b), and the moderating role of participants’ demographic profile (i.e., Age and Gender) on user evaluations across the four outcome themes (H3). Participants were randomly allocated to one of two T/V Distinction experimental conditions in their native language; either to the “T condition” (French: “Tu”; German: “Du”) or to the “V condition” (French: “Vous”; German: “Sie”). Taken together, the experiment corresponded to a 2 (T form vs. V form) x 2 (French vs. German) full-factorial between-groups design. Following the “Checklist for Reporting of Results of Internet E-Surveys” (64), we outline the study design and procedure in detail.

### Procedure and Participants

In total, 284 participants were recruited from the French ( $n = 136$ ) and German ( $n = 148$ ) speaking parts of Switzerland in September 2019. Individuals ranged in Age from 18 to 84 years old ( $M = 41.9$  years,  $SD = 16.6$ ) and 51% were women. Most participants’ highest education attained was a high school diploma (64%), 31% had a university degree. Further details on participant background are available in the **Supplementary Materials**.



**FIGURE 1** | Experimental Stimuli—English Version. Note: “You” form cannot be manipulated in modern English, thus we display one version only.

Participants were recruited *via* Talk Online Panel GmbH, a European specialist research recruitment company. To compensate for their efforts, participants were rewarded based on a points-based incentive system in line with ESOMAR standards. Participants were sent a survey link *via* e-mail by Talk Online, filtered for their native language, and assigned to a translated (German or French) version of the survey accordingly. After answering further screening (>18 years old, native speakers of either German or French) and demographic (i.e., Age, Gender, and Education) questions, participants were randomly allocated to one of the two T/V Distinction experimental conditions in their native language. After interacting with the respective allocated CA prototype, participants were then redirected to complete the rest of the survey with all outcome variables (user evaluation variables and T/V preference) and debriefed as to the experiment’s purpose. In total, the time spent completing the experiment and survey combined ranged from 2.36 mins to 15.15 mins ( $M = 5.65$ ,  $SD = 1.88$ ).

All participants were aware that they could leave the experiment at any time without penalty. Full ethical clearance was given by ETH Zurich Ethics Commission (Ethic’s proposal number: 2019-N-127).

## Development of Experimental Stimuli

The experimental stimuli used were based on a prototype of a healthcare CA developed by a major Swiss health insurance company, created to answer customer queries for health information *via* both text and voice inputs and outputs. Participants could click through the first few conversational turns with the prototype of the CA (named MIA) that was built into a webpage. These conversational turns encompassed the onboarding of the user to the CA interaction rather than the entire medical service. Only the second-person pronouns

used by the CA to address the user were manipulated (i.e., the T/V distinction; T form or V form). For purposes of experiment standardization, the interaction with the CA was purely text-based (i.e., no voice in- or output) and followed a rule-based conversational script with predefined answer options (i.e., graphical buttons). An English translation of the app is depicted in **Figure 1** and the experimental stimuli for each treatment condition are depicted in **Figures 2A–D**. The introductory statement to participants (in English) can be found in the **Supplementary Material**. Manipulation checks confirmed that the experimental conditions functioned as intended.

## Measurement of Outcome Variables

All measurements that focused on measuring attitudes of and perceptions toward the chatbot (to investigate H2a, H2b, and H3) were adapted from established multi-item scales whenever possible (e.g., Social Presence, Trust, etc.; see **Table 1**). Variables from Davis’ (72) Technology Adoption Model were measured based on single-items to reduce the workload for participants as several previous CA studies have used single items for these variables as well (e.g., Liao et al. (73), Oh et al. (74), and Shamekhi et al. (75)). All aforementioned items were measured on 7-point Likert scales ranging from 1 = “Completely disagree” to 7 = “Completely agree.”

The Net Promoter Score, which consists of one item by design (71), was measured on a 10-point Likert scale ranging from 0 = “Very unlikely” to 9 = “Very likely.” Intended Usage of the CA [adapted from Wixom and Todd (66)] was measured on an 11-point Likert scale ranging from 0 = “Very unlikely” to 10 = “Very likely.” Intended Usage Frequency (own scale) was measured on a 7-point Likert scale ranging from 1 = “Never” to 7 = “Very often (several times a day).”

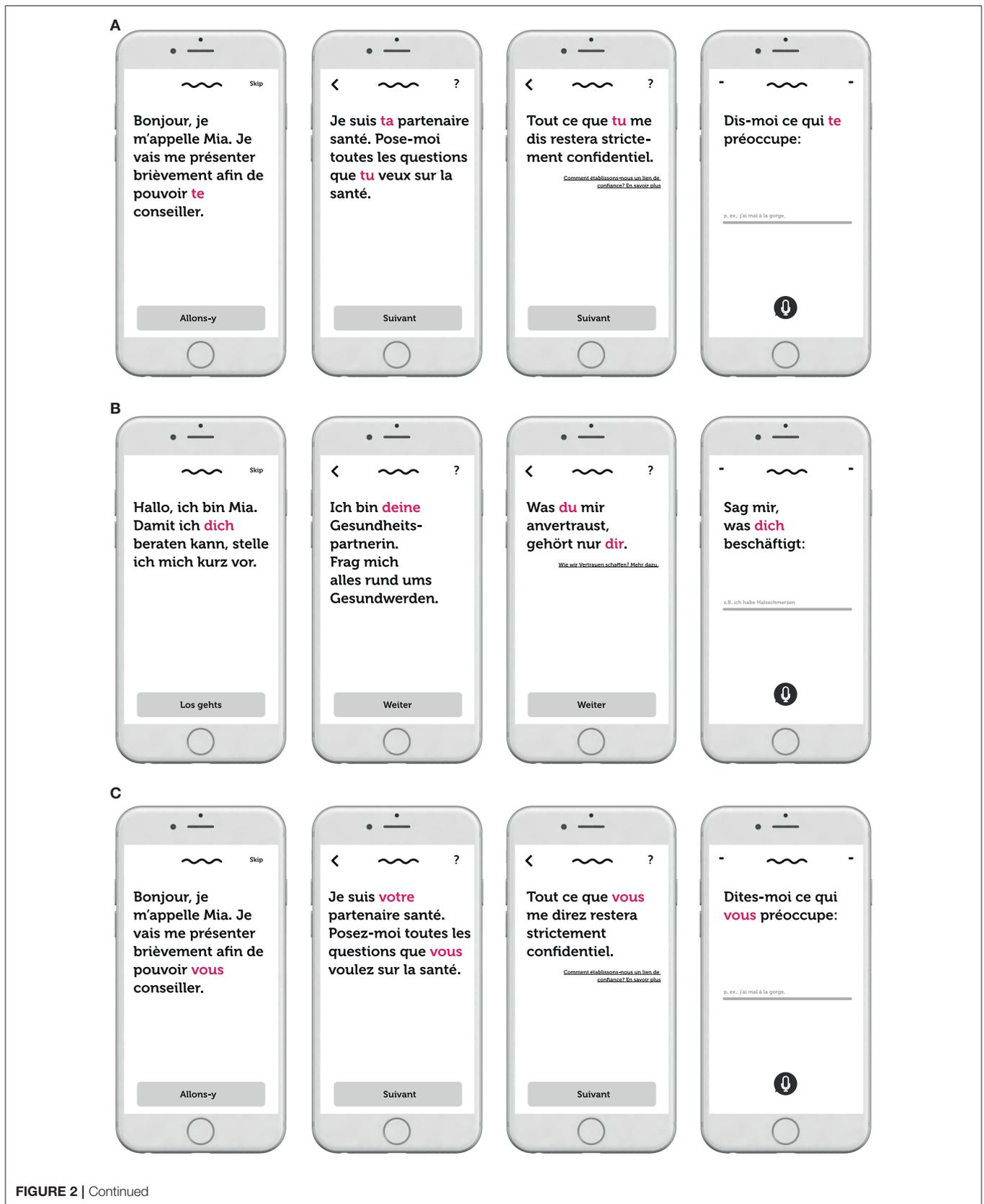
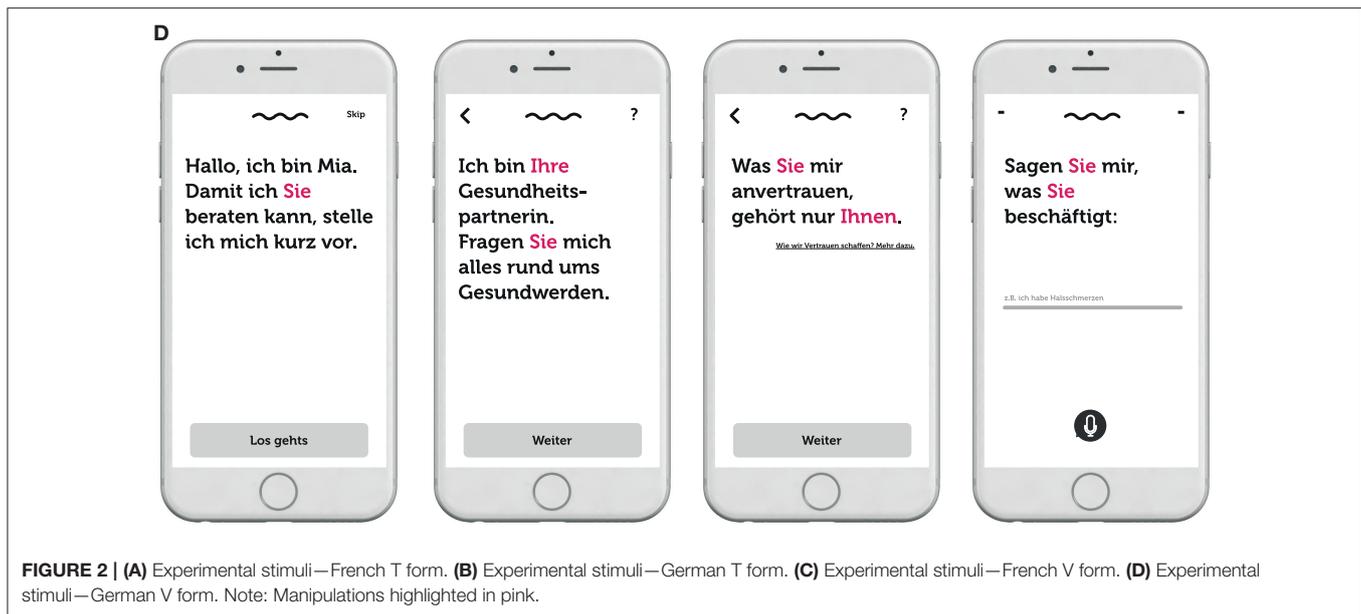


FIGURE 2 | Continued



Addressing H1a and H1b, we also collected a self-created measure (named T/V Preference) of users' rated subjective preference for either the T or V form ("If you had the choice: Would you rather like MIA to use [T form] or [V form] with you?") measured on an 11-point Likert scale ranging from 1 = "[T form]" to 11 = "[V form]."

## Statistical Analysis

Prior to analysis, data were filtered for missing responses, language, and other attention and quality checks, with 284 responses included in the final analyses. Where constructs consisted of multiple items, reliability analyses were carried out to discern Cronbach's alpha with all constructs scoring  $>0.70$  threshold (76). Histograms and Q-Q plots were used to test for Gaussian distributions of the dependent variables so that parametric tests could be utilized. To not reduce data points, user Age was included as a continuous variable rather than dichotomized into a categorical variable (77). As T/V preference represents a theoretically distinct concept and does not belong to the same system of variables as user evaluations, separate models were specified for T/V preference and user evaluation outcomes (78). Data were analyzed using R version 4.0.5

### T/V Preference

In the first model, which investigated users' subjectively stated T/V preference (H1a and H1b), we specified a single ANCOVA model with type III sum of squares with partial eta squared ( $\eta_p^2$ ) indicating the size of the effect. T/V Distinction, user Language, Gender, and Age were included as independent variables, and all main effects and two-, three- and four-way interactions were investigated (79, 80).

### User Evaluations

In the second model, we investigated the objective effect of the experimental conditions on all outcome variables (H2a,

H2b, and H3). To confirm the suitability of outcome variables from the four themes for MANCOVA analysis, a Pearson's correlation table (Table 2) was first calculated to confirm all outcome variables were below a correlation threshold of  $r = 0.90$ , indicating variables were sufficiently correlated for multivariate analysis but did not exhibit perfect multicollinearity (81). Following this, we specified a MANCOVA model with type III sum of squares, with partial eta squared ( $\eta_p^2$ ) and Wilk's Lambda ( $\Lambda$ ) indicating effect size. For the dependent variables, we used the user evaluation variables from all outcome themes, and again specified T/V Distinction, user Language, Gender, and Age as independent variables and investigated main effects and two-, three- and four-way interactions (79, 80).

Where overall significant effects were discerned by the MANCOVA analysis, we followed the procedure outlined in Stevens (82) and Tabachnick and Fidell (81) and utilized identically specified ANCOVA models to confirm whether the significant independent variable(s) found in MANCOVA analysis also held for each independent variable individually (78). Where this was the case, we utilized further Tukey HSD *post-hoc* tests to discern where significant differences between groups existed while controlling for Type I error (81), obtaining slope estimates and confidence intervals. Following guidance outlined by Field, Miles and Field (80), where significant interactions occurred, we investigated the highest-order interactions and not lower-order interactions or main effects (83).

## RESULTS

### ANCOVA Model Results Reveal Significant Main Effect for Language on T/V Preference

The separate ANCOVA model for T/V Preference revealed a significant main effect for Language with a medium effect size

**TABLE 1** | List of measures.

Construct/Item (Origin)	Reference
<p><b>Social Presence</b> (7-point Likert, 1 = fully disagree – 7 = fully agree)</p> <p>I felt a sense of human contact in Mia.</p> <p>I felt a sense of personalness in Mia.</p> <p>I felt a sense of human warmth in Mia.</p> <p>I felt a sense of sociability in Mia.</p> <p>I felt a sense of human sensitivity in Mia.</p>	Adapted from Gefen and Straub (65)
<p><b>Conversational Enjoyment</b> (7-point Likert, 1 = fully disagree – 7 = fully agree)</p> <p>I enjoyed chatting with Mia</p>	Adapted from Wixom and Todd, (66)
<p><b>Conversational Agent Trust</b> (7-point Likert, 1 = fully disagree – 7 = fully agree)</p> <p>I would rely on Mia.</p> <p>I would trust Mia.</p> <p>Mia is honest.</p>	Adapted from Grohmann, 2009 (67)
<p><b>Co-Production</b> (7-point Likert, 1 = fully disagree – 7 = fully agree)</p> <p>I would openly discuss my health situation with Mia to help her find the best solution for me.</p> <p>I would fully cooperate with Mia.</p> <p>I would be willing to do my best to achieve a good outcome with Mia.</p>	Adapted from Mende and van Doorn (68), Böttgen, Schumann and Ates, (69)
<p><b>Privacy Concern</b> (7-point Likert, 1 = fully disagree – 7 = fully agree)</p> <p>Compared with others, I am more sensitive about the way digital services handle my personal information.</p> <p>To me, it is the most important thing to keep my privacy intact.</p> <p>In general, I am very concerned about threats to my personal privacy</p> <p>For me, protecting my privacy is a top priority.</p>	Adapted from Kehr, Kowatsch and Wentzel (70)
<p><b>Perceived Privacy Protection</b> (7-point Likert, 1 = fully disagree – 7 = fully agree)</p> <p>I feel that my privacy would be maintained when interacting with Mia.</p> <p>I think that my privacy is preserved when interacting with Mia.</p> <p>I would be comfortable with the level of privacy I would have when interacting with Mia.</p>	Adapted from Kehr, Kowatsch and Wentzel (70)
<p><b>Technology adoption model constructs</b> (7-point Likert, 1 = fully disagree – 7 = fully agree)</p> <p><b>Presumed ease of use:</b> Overall, the service seems easy to use.</p> <p><b>Presumed usefulness:</b> I believe that Mia would provide a health service that is useful to me.</p> <p><b>Presumed service satisfaction:</b> All things considered, I would be very satisfied with Mia.</p>	Adapted from Wixom and Todd (66)
<p><b>Net promoter score</b> (10-point Likert, 0 = very unlikely – 9 = very likely)</p> <p>How likely is it that you would recommend Mia to a friend or colleague?</p>	Adapted from Reichheld (71)
<p><b>Intention Usage</b> "Assuming Mia was accessible and fully usable for free, ...</p> <p><b>Intention to Use:</b> "... how likely would you be to use the service?"</p> <p>(11-point Likert, 0 = very unlikely – 10 = very likely)</p> <p><b>Intended Usage Frequency:</b> "... how often would you like to use Mia?"</p> <p>(1 = Never, 2 = Very rarely (= less often than 1x per month), 3 = Rarely (= at least 1x per month), 4 = Sometimes (= at least 1x per week), 5 = Often (= at least 1x daily), 6 = Very often (= several times daily)</p>	Adapted from Wixom and Todd (66)
<p><b>T/V Preference</b></p> <p>If you had the choice: Would you prefer Mia to use T- or V-form? (German: Wenn Sie die Wahl hätten: Wäre es Ihnen lieber, wenn Mia Sie „duzen“ oder „siezzen“ würde?)</p> <p>I'd prefer it if Mia used ... 1 = T-form, 7 = V-form</p>	Own scale
<p><b>Socio-/demographics</b></p> <p>What is your <b>gender</b>? (female, male, other); How <b>old</b> are you? (free, text entry); Please indicate your <b>native language</b>. (German, French, Other); What is the highest <b>educational</b> degree you have? (No school-leaving qualification, Elementary school, Secondary school, High school (Matura), Apprenticeship, University, Bachelor), University (Master), Doctorate, Other)</p>	Own scale

All instructions and items were translated to French and German respectively.

**TABLE 2** | Pearson correlations of dependent variables.

	M	SD	1	2	3	4	5	6	7	8	9	10	11
1. CA trust	4.02	1.69	(0.93)										
2. Co-production	4.34	1.74	<b>0.887***</b>	(0.95)									
3. Privacy concern	5.23	1.36	0.120	0.106	(0.88)								
4. Privacy protection	3.95	1.67	<b>0.778***</b>	<b>0.738***</b>	0.008	(0.96)							
5. Social presence	3.88	1.59	<b>0.720***</b>	<b>0.684***</b>	<b>0.142*</b>	<b>0.686***</b>	(0.87)						
6. Convers. enjoyment	4.26	1.68	<b>0.743***</b>	<b>0.757***</b>	0.052	<b>0.675***</b>	<b>0.712***</b>	(-)					
7. Perceived ease of use	5.74	1.29	<b>0.483***</b>	<b>0.513***</b>	0.095	<b>0.403***</b>	<b>0.428***</b>	<b>0.475***</b>	(-)				
8. Perceived usefulness	4.51	1.73	<b>0.788***</b>	<b>0.823***</b>	0.110	<b>0.693***</b>	<b>0.689***</b>	<b>0.816***</b>	<b>0.524***</b>	(-)			
9. Service satisfaction	4.65	1.53	<b>0.718***</b>	<b>0.765***</b>	0.054	<b>0.637***</b>	<b>0.632***</b>	<b>0.781***</b>	<b>0.531***</b>	<b>0.798***</b>	(-)		
10. Intended usage	3.15	1.27	<b>0.704***</b>	<b>0.689***</b>	<b>0.176**</b>	<b>0.581***</b>	<b>0.571***</b>	<b>0.577***</b>	<b>0.252***</b>	<b>0.610***</b>	<b>0.590***</b>	(0.70)	
11. Net promoter score	4.78	2.54	<b>0.842***</b>	<b>0.856***</b>	0.082	<b>0.721***</b>	<b>0.751***</b>	<b>0.821***</b>	<b>0.462***</b>	<b>0.831***</b>	<b>0.791***</b>	<b>0.660***</b>	(-)

\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$  (boldface). In correlation matrix diagonal Cronbach's  $\alpha$  are displayed in brackets.

(83),  $F_{(1,268)} = 14.79$ ,  $p = <0.001$ ,  $\eta_p^2 = 0.05$ , with mean scores indicating that French-speaking participants preferred the formal V form ( $M = 7.18$ ,  $SE = 0.32$ ) compared to German-speaking participants who preferred the informal T form ( $M = 4.72$ ,  $SE = 0.30$ ). No other main or interaction effects were discerned. Taken together, both H1a (i.e., that French speakers will exhibit a preference for the formal “vous”) and H1b (i.e., that German speakers will exhibit a preference for the informal “du”) are fully supported.

## MANCOVA Model Results Reveals Significant Four-Way-Interaction Effect on User Evaluations

The MANCOVA model specified using all outcome variables showed no significant main effects, however, interaction effects were found for T/V Distinction and Gender ( $Wilks' \lambda = 0.924$ ,  $F_{(11,258)} = 1.939$ ,  $p = 0.035$ ,  $\eta_p^2 = 0.076$ ), T/V Distinction, Gender and Age ( $Wilks' \lambda = 0.921$ ,  $F_{(11,258)} = 2.014$ ,  $p = 0.027$ ,  $\eta_p^2 = 0.079$ ), T/V Distinction, Language, and Gender ( $Wilks' \lambda = 0.907$ ,  $F_{(11,258)} = 2.404$ ,  $p = 0.007$ ,  $\eta_p^2 = 0.093$ ), and T/V Distinction, Language, Gender, and Age ( $Wilks' \lambda = 0.903$ ,  $F_{(11,258)} = 2.531$ ,  $p = 0.005$ ,  $\eta_p^2 = 0.097$ ) showing effect sizes ranging from medium to small (83). **Table 3** yields an overview of the MANCOVA model results. Additionally, as the continuous independent variable Age was found significant, we also investigated if multivariate curvilinear trends were present by including a 2nd degree polynomial term for Age in the above model, however, no significant improvement to model fit was found (see **Supplementary Materials**), confirming suitability for linear analyses.

## Follow-Up ANCOVA Models

As the MANCOVA discerned a significant four-way interaction between T/V Distinction, Language, Gender, and Age, ANCOVA models for each outcome variable were specified in the same manner (78, 81, 82). In each of the ANCOVA models across all outcome variables, the significant four-way interaction was again confirmed with the exceptions of Privacy Concern (PC) ( $p = 0.120$ ) and Perceived Ease of Use (PEOU) ( $p = 0.400$ ) which

were insignificant and therefore not further analyzed in the *post-hoc* analyses. An excerpt of the ANCOVA models results for the four-way interaction effect can be found in **Table 4**.

## Post-hoc Tests Reveal Consistent Four-Way Interaction Pattern for Nine Out of 11 Outcome Variables

To further explore the four-way interactions confirmed in the ANCOVA models, subsequent *post-hoc* Tukey HSD tests were conducted to estimate the overall slope coefficients and confidence intervals for T/V Distinction, Language, Gender by Age; graphically represented in **Figures 3A–D** and summarized in **Tables 5A–D**. The results showed a four-way interaction between T/V Distinction, Gender, Language, and Age for all included user evaluation variables with a consistent pattern across all outcome themes. The pattern evidenced is described in the following passages using the outcomes of Net Promoter Score (NPS) and Social Presence as examples:

For French speakers, when the informal T form (“Tu”) was used, higher user evaluation scores were generated for younger women and older men, whereas when the formal V form (“Vous”) was used, higher user evaluation scores were generated for younger men and older women respectively. For example, for French women in the T condition, as their Age increased, Net Promoter Score ( $\beta = -0.067$ ,  $SE = 0.028$ ) and Social Presence ( $\beta = -0.036$ ,  $SE = 0.018$ ) scored lower (i.e., individuals rated lower likelihood to recommend MIA to friends or relatives and MIA felt less humanlike) whereas French women in the V condition scored higher (i.e., individuals rated higher likelihood to recommend MIA to friends or relatives and MIA felt more humanlike) with Net Promoter Score ( $\beta = 0.068$ ,  $SE = 0.028$ ) and Social Presence ( $\beta = 0.031$ ,  $SE = 0.018$ ). Conversely, for French men in the T condition, as their Age increased, Net Promoter Score ( $\beta = 0.006$ ,  $SE = 0.025$ ) and Social Presence ( $\beta = 0.011$ ,  $SE = 0.015$ ) rated higher (i.e., individuals rated higher likelihood to recommend MIA to friends or relatives and MIA felt more humanlike), whereas French men in the V condition scored lower (i.e., individuals rated lower likelihood to recommend MIA

**TABLE 3** | MANCOVA model results.

MANCOVA	df	Wilks' $\lambda$	F	p	$\eta_p^2$
<b>Main effects</b>					
T/V distinction	11	0.962	0.917	0.524	0.038
Language	11	0.940	1.500	0.131	0.060
Age	11	0.935	1.622	0.093	0.065
Gender	11	0.972	0.655	0.781	0.021
<b>Two-way interaction effects</b>					
T/V distinction * language	11	0.978	0.518	0.891	0.022
T/V distinction * age	11	0.964	0.872	0.569	0.036
T/V distinction * gender	<b>11</b>	<b>0.924</b>	<b>1.939</b>	<b>0.035</b>	<b>0.076</b>
Language * age	11	0.955	1.102	0.360	0.045
Language * gender	11	0.970	0.732	0.708	0.030
Gender * age	11	0.979	0.503	0.900	0.021
<b>Three-way interaction effects</b>					
T/V distinction * language * age	11	0.975	0.601	0.828	0.025
T/V distinction * language * gender	<b>11</b>	<b>0.907</b>	<b>2.404</b>	<b>0.007</b>	<b>0.093</b>
T/V distinction * gender * age	<b>11</b>	<b>0.921</b>	<b>2.014</b>	<b>0.027</b>	<b>0.079</b>
Language * gender * age	11	0.980	0.481	0.914	0.020
<b>Four-way interaction effects</b>					
T/V Distinction * language * gender * age	<b>11</b>	<b>0.903</b>	<b>2.531</b>	<b>0.005</b>	<b>0.097</b>

Significant values at  $p < 0.05$  are in boldface.

**TABLE 4** | ANCOVA models results for four-way interaction between T/V Distinction, language, gender, and age per outcome variable.

Theme	Dependent variable	F	p	$\eta_p^2$
<b>Sociability</b>	Social presence	<b>8.447</b>	<b>0.004</b>	<b>0.031</b>
	Conversational enjoyment	<b>5.769</b>	<b>0.017</b>	<b>0.021</b>
<b>CA-User Collaboration</b>	Conversational agent trust	<b>4.667</b>	<b>0.032</b>	<b>0.017</b>
	Privacy concern	2.421	0.120	0.009
	Perceived privacy protection	<b>6.731</b>	<b>0.010</b>	<b>0.024</b>
	Co-production	<b>7.997</b>	<b>0.005</b>	<b>0.029</b>
<b>Service Evaluation</b>	Perceived ease of use	0.710	0.400	0.003
	Perceived usefulness	<b>12.480</b>	<b>0.000</b>	<b>0.044</b>
	Service satisfaction	<b>4.728</b>	<b>0.031</b>	<b>0.017</b>
<b>Behavioral Intentions</b>	Intended usage	<b>9.295</b>	<b>0.003</b>	<b>0.034</b>
	Net Promoter score	<b>10.923</b>	<b>0.001</b>	<b>0.039</b>

Significant values at  $p < 0.05$  are in boldface. Only four-way interaction results are displayed.

to friends or relatives and MIA felt less humanlike) with Net Promoter Score ( $\beta = -0.055$ ,  $SE = 0.022$ ) and Social Presence ( $\beta = -0.033$ ,  $SE = 0.017$ ) respectively.

For German speakers, when the informal T form (“Du”) was used, younger users’ evaluation scores rated comparably regardless of Gender, however, as individuals’ Age increased, the use of “Du” resulted in relatively lower user evaluation scores, and this effect was even more pronounced for men. Whereas, in the formal V condition (“Sie”), user evaluation scores were relatively stable, regardless of Gender, and showed only a slight

influence of Age. For example, in the informal T condition as users’ Age increased, Net Promoter Score ( $\beta = 0.027$ ,  $SE = 0.024$ ) and Social Presence ( $\beta = 0.028$ ,  $SE = 0.015$ ) rated lower for women, and even lower for men (i.e., individuals rated lower likelihood to recommend MIA to friends or relatives and MIA felt less humanlike) with Net Promoter Score ( $\beta = 0.069$ ,  $SE = 0.029$ ) and Social Presence ( $\beta = 0.047$ ,  $SE = 0.018$ ) respectively. Whereas in the formal V condition trends were comparably stable between Genders as users’ Age increased [e.g., Net Promoter Score: male ( $\beta = 0.021$ ,  $SE = 0.022$ ) vs. female ( $\beta$

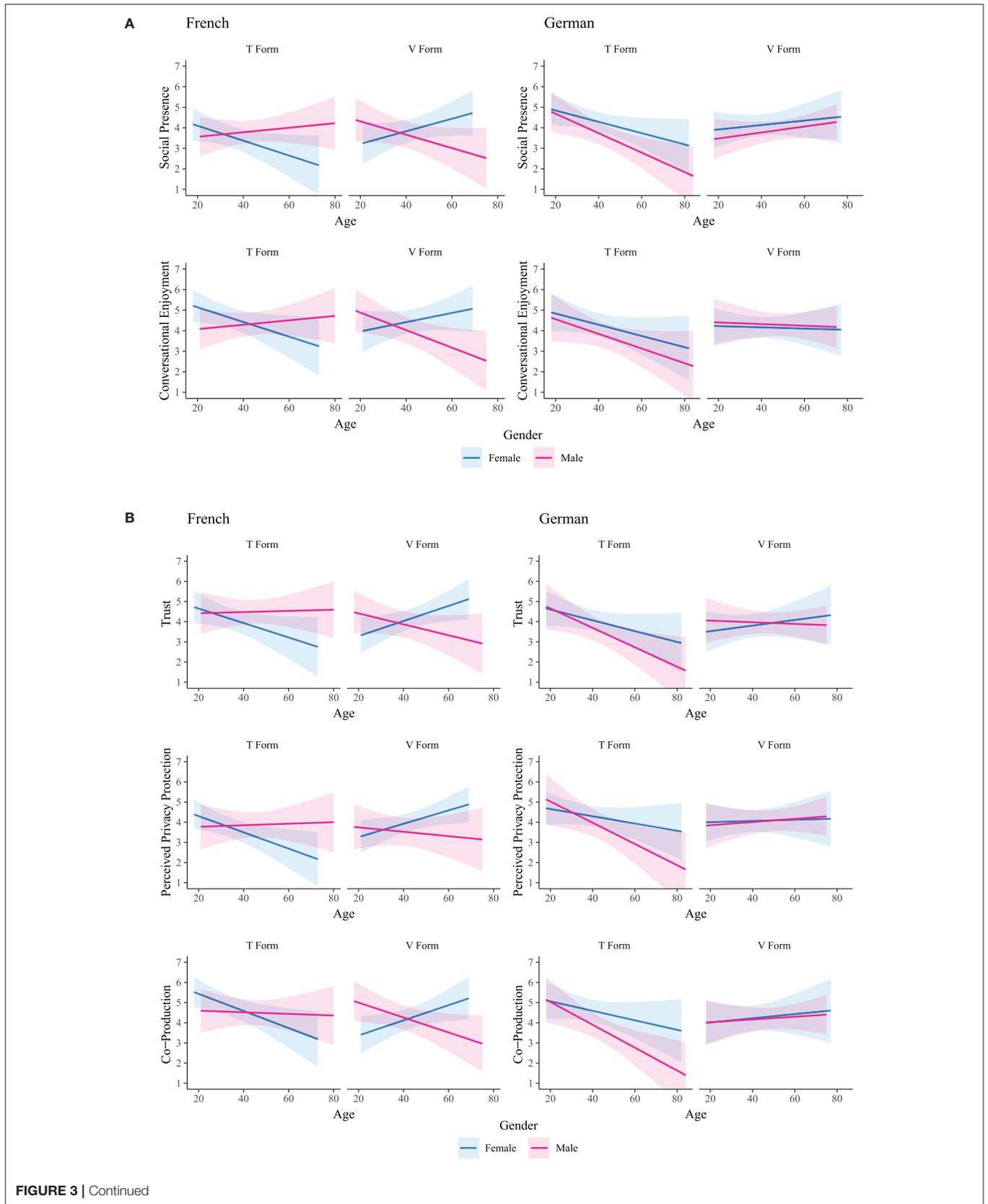
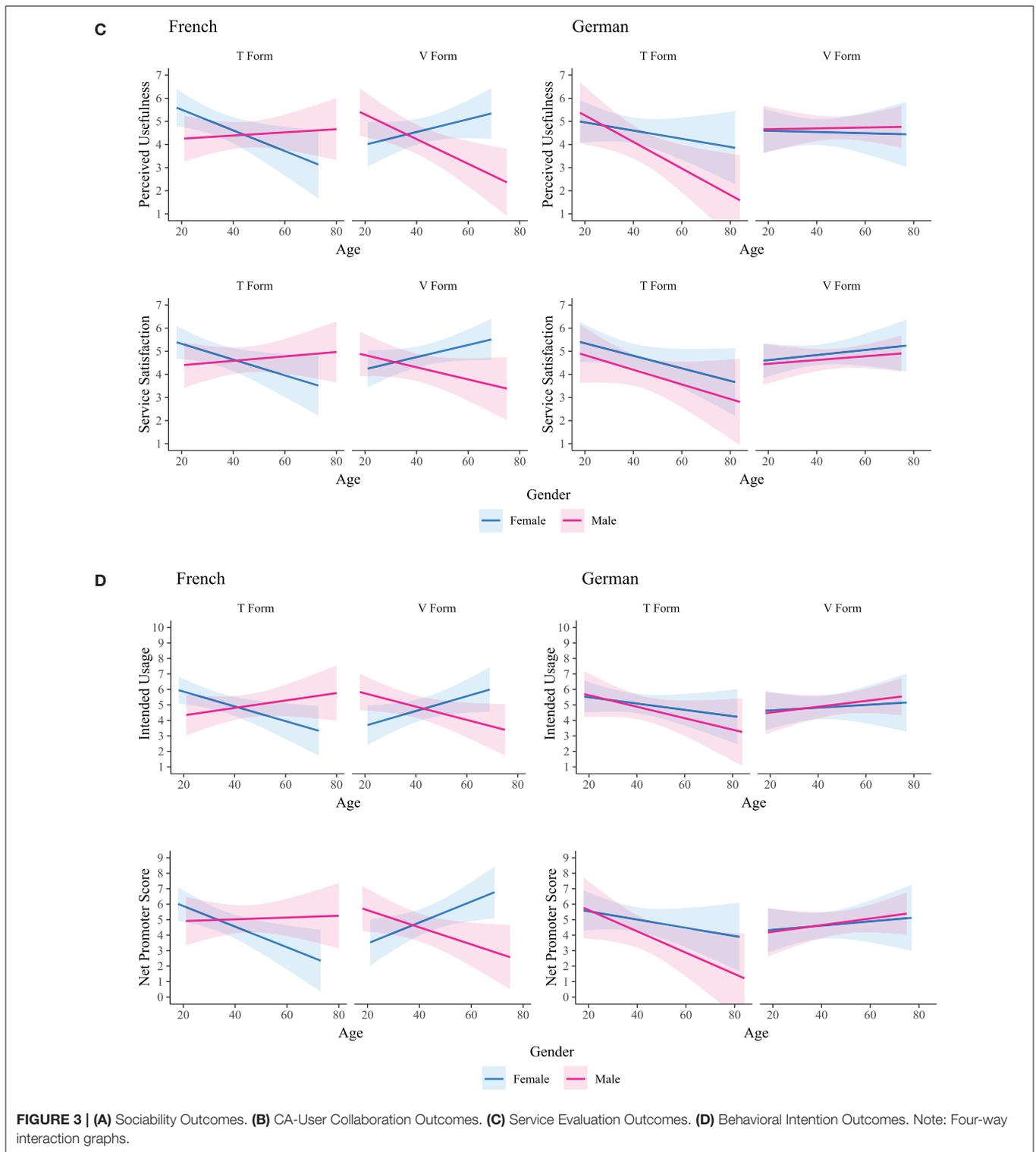


FIGURE 3 | Continued



= 0.014, SE = 0.026) scores; Social Presence: male ( $\beta = 0.015$ , SE = 0.014) vs. female ( $\beta = 0.011$ , SE = 0.016)] scores (i.e., for both genders as individuals' age increased, individuals stated a slight increased likelihood to recommend MIA to friends or relatives and rated MIA more humanlike).

### Pairwise Comparisons

Further Tukey pairwise comparisons were conducted with a summary of the significant differences in the slope trend between groups (as calculated by subtracting comparison group,  $\beta_c$ , from the reference group,  $\beta_r$ ) summarized in **Table 6**. Findings

TABLE 5 | Slope coefficients.

Measure	T/V	Language	Gender	Age Slope	SE	Lower CI	Upper CI
<b>(A)</b>							
<b>(i) Sociability</b>							
<i>Social Presence</i>	T	FR	F	-0.036	0.018	-0.071	-0.002
	V	FR	F	0.031	0.018	-0.004	0.065
	T	GER	F	-0.028	0.015	-0.057	0.002
	V	GER	F	0.011	0.016	-0.020	0.042
	T	FR	M	0.011	0.015	-0.019	0.041
	V	FR	M	-0.033	0.017	-0.066	0.000
	T	GER	M	-0.047	0.018	-0.083	-0.012
	V	GER	M	0.015	0.014	-0.013	0.042
<i>Conversational Enjoyment</i>	T	FR	F	-0.036	0.019	-0.073	0.001
	V	FR	F	0.023	0.019	-0.015	0.060
	T	GER	F	-0.027	0.016	-0.060	0.005
	V	GER	F	-0.003	0.017	-0.037	0.031
	T	FR	M	0.011	0.017	-0.022	0.043
	V	FR	M	-0.043	0.018	-0.078	-0.007
	T	GER	M	-0.036	0.019	-0.074	0.002
	V	GER	M	-0.004	0.015	-0.034	0.025
<b>(B)</b>							
<b>(ii) CA-User Collaboration</b>							
<i>Trust in CA</i>	T	FR	F	-0.036	0.019	-0.073	0.001
	V	FR	F	0.037	0.019	0.000	0.075
	T	GER	F	-0.027	0.016	-0.059	0.005
	V	GER	F	0.014	0.017	-0.020	0.047
	T	FR	M	0.003	0.017	-0.030	0.035
	V	FR	M	-0.027	0.018	-0.063	0.008
	T	GER	M	-0.048	0.019	-0.086	-0.011
	V	GER	M	-0.004	0.015	-0.034	0.025
<i>Perceived Privacy Protection</i>	T	FR	F	-0.040	0.019	-0.077	-0.004
	V	FR	F	0.033	0.019	-0.004	0.070
	T	GER	F	-0.018	0.016	-0.050	0.014
	V	GER	F	0.003	0.017	-0.030	0.036
	T	FR	M	0.004	0.017	-0.029	0.036
	V	FR	M	-0.011	0.018	-0.046	0.024
	T	GER	M	-0.053	0.019	-0.090	-0.015
	V	GER	M	0.008	0.015	-0.021	0.037
<i>Co-Production Behavior</i>	T	FR	F	-0.042	0.019	-0.080	-0.005
	V	FR	F	0.038	0.019	0.000	0.075
	T	GER	F	-0.024	0.017	-0.056	0.009
	V	GER	F	0.010	0.017	-0.024	0.045
	T	FR	M	-0.004	0.017	-0.037	0.029
	V	FR	M	-0.037	0.018	-0.073	-0.001
	T	GER	M	-0.057	0.020	-0.095	-0.018
	V	GER	M	0.007	0.015	-0.023	0.037
<b>(C)</b>							
<b>(iii) Service evaluation</b>							
<i>Perceived Usefulness</i>	T	FR	F	-0.045	0.019	-0.082	-0.007
	V	FR	F	0.028	0.019	-0.010	0.065
	T	GER	F	-0.018	0.017	-0.050	0.015
	V	GER	F	-0.003	0.017	-0.037	0.031

(Continued)

**TABLE 5 |** Continued

Measure	T/V	Language	Gender	Age Slope	SE	Lower CI	Upper CI
	T	FR	M	0.007	0.017	-0.026	0.040
	V	FR	M	-0.053	0.018	-0.089	-0.018
	T	GER	M	-0.057	0.019	-0.096	-0.019
	V	GER	M	0.002	0.015	-0.028	0.032
<i>Service Satisfaction</i>	T	FR	F	-0.034	0.017	-0.068	0.000
	V	FR	F	0.026	0.017	-0.008	0.060
	T	GER	F	-0.027	0.015	-0.056	0.002
	V	GER	F	0.011	0.016	-0.020	0.042
	T	FR	M	0.010	0.015	-0.020	0.039
	V	FR	M	-0.026	0.016	-0.059	0.006
	T	GER	M	-0.032	0.018	-0.066	0.003
	V	GER	M	0.008	0.014	-0.019	0.035
<b>(D)</b> <b>(iv) Behavioral Intentions</b> <i>Intended Usage</i>	T	FR	F	-0.025	0.014	-0.053	0.004
	V	FR	F	0.023	0.014	-0.005	0.051
	T	GER	F	-0.001	0.013	-0.026	0.023
	V	GER	F	0.007	0.013	-0.018	0.033
	T	FR	M	0.025	0.013	0.000	0.050
	V	FR	M	-0.017	0.014	-0.044	0.010
	T	GER	M	-0.021	0.015	-0.050	0.007
	V	GER	M	0.013	0.011	-0.009	0.036
<i>Net Promoter Score</i>	T	FR	F	-0.067	0.028	-0.122	-0.011
	V	FR	F	0.068	0.028	0.012	0.123
	T	GER	F	-0.027	0.024	-0.075	0.021
	V	GER	F	0.014	0.026	-0.037	0.064
	T	FR	M	0.006	0.025	-0.043	0.054
	V	FR	M	-0.055	0.027	-0.108	-0.002
	T	GER	M	-0.069	0.029	-0.125	-0.012
	V	GER	M	0.021	0.022	-0.023	0.065

**TABLE 6 |** Tukey HSD pairwise comparisons.

DV	Reference group	Comparison group(s)	$\beta_r - \beta_c$	SE	p
Conversational Agent Trust	V, FR, F	T, GER, M	<b>0.086</b>	<b>0.027</b>	<b>0.034</b>
Perceived Privacy Protection	V, FR, F	T, GER, M	<b>0.086</b>	<b>0.027</b>	<b>0.032</b>
Co-Production	V, FR, F	T, GER, M	<b>0.094</b>	<b>0.027</b>	<b>0.015</b>
	V, FR, F	V, FR, M	0.075	0.027	0.097
	V, FR, F	T, FR, F	0.080	0.027	0.067
Social Presence	V, FR, F	T, GER, M	<b>0.078</b>	<b>0.025</b>	<b>0.042</b>
Perceived Usefulness	V, FR, F	T, GER, M	<b>0.085</b>	<b>0.027</b>	<b>0.041</b>
	V, FR, F	V, FR, M	<b>0.081</b>	<b>0.026</b>	<b>0.047</b>
Net Promoter Score	V, FR, F	T, GER, M	<b>0.136</b>	<b>0.039</b>	<b>0.018</b>
	V, FR, F	V, FR, M	<b>0.123</b>	<b>0.039</b>	<b>0.038</b>
	V, FR, F	T, FR, F	<b>0.134</b>	<b>0.040</b>	<b>0.019</b>

Significant values at  $p < 0.05$  are in boldface.

outlined that as individuals' Age increased, French-speaking women in the V condition exhibited significantly higher user evaluation scores than German-speaking men in the T condition for Trust ( $\beta_r - \beta_c = 0.086$ ,  $SE = 0.027$ ,  $p = 0.034$ ), Perceived Privacy Protection ( $\beta_r - \beta_c = 0.086$ ,  $SE = 0.027$ ,  $p = 0.032$ ),

Co-Production ( $\beta_r - \beta_c = 0.094$ ,  $SE = 0.027$ ,  $p = 0.015$ ), Social Presence ( $\beta_r - \beta_c = 0.078$ ,  $SE = 0.025$ ,  $p = 0.042$ ), Perceived Usefulness ( $\beta_r - \beta_c = 0.085$ ,  $SE = 0.027$ ,  $p = 0.041$ ) and Net Promoter Score ( $\beta_r - \beta_c = 0.136$ ,  $SE = 0.039$ ,  $p = 0.018$ ). Further significant differences were found in a similar pattern with

French-speaking women in the V condition rating Net Promoter Score ( $\beta_r - \beta_c = 0.123$ ,  $SE = 0.039$ ,  $p = 0.038$ ) and Perceived Usefulness ( $\beta_r - \beta_c = 0.081$ ,  $SE = 0.026$ ,  $p = 0.047$ ) higher as their Age increased than French-speaking men in the V condition. For Net Promoter Score ( $\beta_r - \beta_c = 0.134$ ,  $SE = 0.040$ ,  $p = 0.019$ ), French-speaking women in the V condition also significantly differed from French-speaking women in the T condition as their Age increased. Other marginally significant ( $p < 0.1$ ) pairwise comparisons were also evidenced for Co-Production.

Taken together, we find partial support for H2a (i.e., that CA use of formal "vous" with French-speaking participants causes higher user evaluation scores) as this occurred only for certain user groups (younger men, older women). Similarly, we find partial support for H2b (i.e., that CA use of informal "du" with German-speaking participants causes higher user evaluation scores) as this again occurred only for certain user groups (younger men, younger women). Additionally, we can fully confirm H3 (i.e., that user Age and Gender jointly moderate the relationship between T/V distinction and Language and user evaluation scores) as in each language setting there was not a single T/V form that caused highest user evaluation outcomes, rather, highest user evaluation scores depend on both the Language used and the users' demographic profile (i.e., Age, Gender). The results, therefore, confirm the importance of both linguaculture *per se* (with differences evidenced between both French and German speakers) as well as the importance of demographic profiling of users within a linguaculture (with differences evidenced by Gender and Age).

## DISCUSSION

### Theoretical Contributions

In this study, across all four user evaluation outcome themes, we have shown that the term of address (T/V distinction) employed by CAs varies in suitability both between linguacultures (i.e., French, German) but also within linguacultures by demographic profiling (i.e., Gender, Age). To the authors' knowledge, this is the first time T/V distinction has been linked to a wide range of CA-relevant outcomes, making three main theoretical contributions relevant to the design of CA-based digital health interventions.

First, regarding the outcome themes [i.e., (i) Sociability, (ii) CA-User Collaboration, (iii) Service Evaluation, and (iv) Behavioral Intentions], we demonstrate how the term of address has a wide-reaching impact on a variety of user evaluation outcomes. Findings from the (i) Sociability theme underscore the importance of anthropomorphism (18, 84) and our research begins the process of linking specific linguistic cues to perceptions of humanness. As considerable angst surrounds T/V distinction usage (52), it is likely that when presented with the T/V form that most closely matches socially-based expectations, MIA (the CA) was perceived as a more efficacious and socially experienced dyadic partner and thus higher ratings of Social Presence and Conversational Enjoyment were found. For the Service Evaluation theme, this may explain why Perceived Usefulness and Service Satisfaction were significant while Perceived Ease of Use was not: Individuals likely evaluated

the former two variables in terms of their relationship with MIA (the CA) as a dyadic partner, whereas the latter was based on the app interface. Additionally, as enabling working relationships between users and healthcare CAs has been a widely desired outcome (15), findings for the (ii) CA-User Collaboration theme show how appropriate use of T/V form facilitates CA-user collaboration by increasing Trust, Perceived Privacy Protection, and Co-Production (i.e., ability to work well with the CA). It may be that for those groups where T form caused higher user evaluation scores, its use created a greater sense of CA-User relatedness (48) facilitating feelings of trust and security, whereas, for those groups where V form caused higher user evaluation scores, its usage conveys themes of power asymmetry (48) and thus connotations of formality, professionalism, or respect.

Second, our findings highlight how evaluative processes vary in a distinct manner between user groups. Pairwise comparisons revealed that for older French-speaking women the use of V form was something relatively value-adding (causing improvements in user evaluation scores), whereas for older German-speaking men the lack of V form use (i.e., using T form) was value-reducing (causing reductions in user evaluation scores). As memory is organized in associative networks (85), it may be that semantic cues in T or V forms convey different levels of interpersonal closeness, formality, or professionalism (48), which serve to confirm or disconfirm expectations for different user groups (51). Previous research has shown, that age, social status, relationships (friends, colleagues, acquaintances) (50, 53), and physical location of conversation (52) are all part of an evaluative process influencing the selection of a cultural script (44), and in turn, reception of linguistic devices used (45) such as terms of address (41). We posit therefore that CA use of T/V distinction confirms the appropriate cultural script for users (based on some underlying implicit need) and facilitates improved user evaluations.

Third, regarding users' subjectively stated T/V preference, we reinforce prior findings showing that preferences vary between linguacultures (41, 49), as French speakers rated a significantly higher preference for the V form. This is likely due to the V form being relatively more common in French in a variety of usage contexts in everyday life, whereas in German it is used less frequently and typically in professional settings (49). Surprisingly, Age was not significantly linked to T/V Preference, in contrast to previous findings (50), which could be taken as evidence of the globalization effect on T/V distinction (43). We would however posit an alternative reasoning: Results from our experiment showed contradictions between users' subjectively stated preferences and objective experimental effects generating best user evaluations. For example, despite French-speaking users stating a preference for V form and Age having no significant influence on T/V preference, results showed that, for older French-speaking men, highest user evaluation scores were generated by T form. Thus, despite T/V Preference not significantly differing by users' Age and Gender, we can see that T/V Distinction usage nonetheless has important effects on user evaluation outcomes, moderated by participants' Age and Gender, in the two language settings. This contradiction in stated preferences in healthcare is known as "hypothetical bias,"

whereby individuals' reported preferences are not congruent to outcomes in a real situation (86, 87) thus underlining the importance of using experimental research when designing healthcare CAs.

## Managerial Implications

The present research offers the following managerially relevant contributions:

First, we demonstrate that unique linguistic-cultural features are still of relevance in the era of globalization. This is of importance as an assumption in some practice-led circles has been that greater globalization will lead to the homogenization of preferences between nations (43), and in healthcare, this is often reflected in standardized approaches when transferring knowledge across borders (88). Indeed, in pursuit of scaling-up health interventions, rapid implementation is often encouraged to address the digital divide between developed and developing nations (89), whilst at the same time, greater personalization of healthcare is known to be both more beneficial as well as a strength readily delivered by digital tools such as CAs (13). While not wishing to dissuade the important rollout of new technologies, our research suggests that considering unique linguistic-cultural features within a linguaculture (such as T/V distinction) and optimizing CA dialogues accordingly would be a worthwhile step: Particularly as relevant demographic user characteristics can be readily elicited early in dialogues with the CA and subsequently utilized for personalization (90).

Second, our findings show that the term of address used significantly and robustly affects managerially relevant outcomes from the Behavioral Intentions theme. Pairwise comparisons revealed a significant effect of CA use of T/V distinction on the Net Promoter Score, a construct strongly linked to commercial success (91), that is also becoming more widely used in healthcare (12, 92). Additionally, pairwise comparisons revealed a significant effect of address form on participants' Intention to Use the CA in the future. Although intentions may not (always) lead to behavior (93), forging positive attitudes, associations, and intentions are still important steps in behavior change (94), helping to create engaging user experiences that patients adhere to, which is, ultimately, vital for treatment success (95). Our results, therefore, demonstrate a practical method by which practitioners can improve peer-network recommendations and adoption of their CA-based digital health interventions.

In summary, while maintaining long-term CA-user relationships will likely vary on many factors, T/V distinction and other linguistic devices may have a greater role in first impression management, and potentially also longer-term interactions too, than previously realized.

## Limitations and Future Research

Although greatly elucidating how terms of address used by CAs affect user evaluations, further research can build upon our findings in several research directions (RDs). The current research examined two unique linguacultures of French and German speakers, however, within a single multi-lingual nation of Switzerland. Switzerland's unique context allowed us to mitigate potential bias from other inter-national/cultural

differences, for example, between different nation-states (96, 97). However, users from Switzerland may be more culturally homogenous than say French speakers from France and German speakers from Germany, which may have caused less variation in user evaluations. Additionally, while no dialect was employed in the app (i.e., standard French and German was used), it could be that individuals from non-geographically proximal locations to Switzerland that historically have used different dialects (for example, French-speakers in Belgium or German-speakers in northern Germany) may exhibit some variation in evaluations. Further research should therefore examine users from two distinct nations, which may confirm trends identified (RD1).

Another consideration would be a comparison of effect sizes. The current research discerned a medium effect size of Language on (subjectively stated) T/V Preference and a small effect size of (objectively manipulated/controlled) T/V distinction, Language, Gender, and Age on user evaluations. Although previous CA-based research has highlighted that use of users' first names (as the term of address) causes significantly improved CA evaluations, no effect size was reported by the authors (25). Additionally, papers on terms of address in linguistics fields are typically qualitative, using methodologies such as Natural Semantic Metalanguage (49) or ethnographic observation (52). Thus, comparisons of our results with published research are difficult. Some studies including effect sizes of language-based cues used by CAs are available in broader research outside terms of address, for example, Schuetzler, Grimes, and Giboney (1) examined tailored vs. generic communication styles and found tailored communication significantly improved perceived humanness, with a large effect size. Additionally, Rietz et al. (98) found CA use of social cues in dialogues (use of emojis, short pauses) significantly improved ratings of usability outcomes, with small effect sizes. As CA-research is in its infancy, we would therefore encourage future researchers to publish details on experimental manipulations used and effect sizes discerned. In such a way, a taxonomy of linguistic features can be built (RD2) acting as a research reference and practical guide for CA designers. A next step in building this taxonomy, following the current paper, could be to examine T/V distinction (or related facets of speech etiquette) in non-Latin languages (RD3). If findings from the current study were also confirmed in other linguacultures, global universals in CA design could be established, which may enable faster adaptation of health interventions internationally. This would fit with recent calls from the World Health Organization to scale-up CAs (14), and provide an interesting research opportunity to investigate CA speech etiquette in linguacultures around the globe.

An additional consideration regards temporal aspects of the healthcare CA used in the current experiment. Individuals using MIA were interacting for the first time, and a variety of more complex rituals govern first impression management (33). Research from CA literature has shown that long-term relationships typically have more social-emotional components to them (33), which is particularly the case during long-term CA-led interventions to address chronic diseases (22). Therefore,

whilst we can be confident how to make a positive first impression based on the current results, which are particularly applicable to health services such as symptom checking (3), CA T/V distinction appropriateness may change over time for longitudinal interventions, mirroring aspects of relationship development in non-digital contexts (52). Further research may therefore wish to investigate longitudinal aspects of CA term of address use (RD4). In a similar vein, while our results confirmed the importance of Age as a moderator, our results cannot definitively state whether this is due to generational differences that will dissipate over time (50), or whether as individuals age they are socialized into utilizing certain terms of address (63). Moving forward, it will therefore be fascinating to examine how humans and machines address each other, with the advent of newer technologies and longitudinal data (RD5).

Lastly, for certain variables (PC, PEOU), trends were identified that appear to similarly fit the pattern identified for other outcome variables, however, were not statistically significant. This may be because the current user experience with MIA (the CA) was too brief, not realistic enough, or the current sample size was not large enough to detect effects. Future research may wish to examine these variables again with a fully operational prototype (RD6) or larger sample size (RD7).

## CONCLUSION

Conversational agents are driven by the ability to communicate effectively with users, in ways that connect with them and build working alliances. In a variety of linguacultures globally, T/V distinction represents an ever-salient term of address and an important linguistic device. By conforming to users' expectations regarding facets of speech-etiquette and language-use more generally, designers of CAs can shape engaging user experiences. For healthcare CAs this is particularly vital, as it facilitates greater adherence to the treatment plans they help deliver. The current paper, therefore, contributes to a greater understanding of CA linguistics from an international perspective, as well as providing practical steps to design and adapt digital health interventions

## REFERENCES

- Schuetzler RM, Grimes GM, Scott Giboney J. The impact of chatbot conversational skill on engagement and perceived humanness. *J Manag Inf Syst.* (2020) 37:875–900. doi: 10.1080/07421222.2020.1790204
- Kostkova P. Grand challenges in digital health. *Front Public Heal.* (2015) 3:134. doi: 10.3389/fpubh.2015.00134
- Cirković A. Evaluation of four artificial intelligence-assisted self-diagnosis apps on three diagnoses: two-year follow-up study. *J Med Internet Res.* (2020) 22:e18097. doi: 10.2196/18097
- Kretzschmar K, Tyroll H, Pavarini G, Manzini A, Singh I, NeurOx Young People's Advisory G. Can your phone be your therapist? young people's ethical perspectives on the use of fully automated conversational agents (chatbots) in mental health support. *Biomed Inf Insights.* (2019) 11:117822261982908. doi: 10.1177/1178222619829083

cross-culturally: Ultimately aspiring to begin addressing the digital divide between developing and developed nations by increasing the effectiveness of CA use globally.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by ETH Zurich (Ethics proposal number: 2019-N-127). The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

JO and MN were equally responsible for conceptualization, methodology, data curation, and drafting the manuscript. JO was responsible for data analysis. MN for project administration, and funding acquisition. All authors reviewed and approved the manuscript before submission.

## FUNDING

Funding for recruiting participants within the MIA project was provided by CSS insurance.

## ACKNOWLEDGMENTS

The authors would like to thank CSS insurance for their financial support of the study.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpubh.2021.691595/full#supplementary-material>

- Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (woebot): a randomized controlled trial. *JMIR Ment Heal.* (2017) 4:e19. doi: 10.2196/mental.7785
- Alison, Daniels J, Salinger D, Wicks P, Robinson A. Evidence of human-level bonds established with a digital conversational agent: cross-sectional, retrospective observational study. *JMIR Form Res.* (2021) 5:e27868. doi: 10.2196/27868
- Becky, Sarda S, Subramanian V. An empathy-driven, conversational artificial intelligence agent (wysa) for digital mental well-being: real-world data evaluation mixed-methods study. *JMIR Mhealth Uhealth.* (2018) 6:e12106. doi: 10.2196/12106
- Russell, Joerin A, Gentile B, Lakerink L, Rauws M. Using psychological artificial intelligence (tess) to relieve symptoms of depression and anxiety: randomized controlled trial. *JMIR Ment Heal.* (2018) 5:e9782. doi: 10.2196/mental.9782

9. Stephens TN, Joerin A, Rauws M, Werk LN. Feasibility of pediatric obesity and prediabetes treatment support through Tess, the AI behavioral coaching chatbot. *Transl Behav Med.* (2019) 9:440–7. doi: 10.1093/tbm/ibz043
10. Joerin A, Rauws M, Ackerman ML. Psychological artificial intelligence service, tess: delivering on-demand support to patients and their caregivers: technical report. *Cureus.* (2019) 11:e3972. doi: 10.7759/cureus.3972
11. Cottrell E, Chambers R, O’Connell P. Using simple telehealth in primary care to reduce blood pressure: a service evaluation. *BMJ Open.* (2012) 2:e001391. doi: 10.1136/bmjopen-2012-001391
12. Stein N, Brooks K. A fully automated conversational artificial intelligence for weight loss: longitudinal observational study among overweight and obese adults. *JMIR Diabetes.* [2017] 2:e28. doi: 10.2196/diabetes.8590
13. Tudor Car L, Dhinakaran DA, Kyaw BM, Kowatsch T, Joty S, Theng YL, et al. Conversational agents in health care: scoping review and conceptual analysis. *J Med Internet Res.* (2020) 22:e17158. doi: 10.2196/17158
14. Luxton DD. Ethical implications of conversational agents in global public health. bulletin of the world health organization. *World Health Organ.* (2020) 20:285–7. doi: 10.2471/BLT.19.237636
15. Bickmore T, Gruber A, Picard R. Establishing the computer-patient working alliance in automated health behavior change interventions. *Patient Educ Couns.* (2005) 59:21–30. doi: 10.1016/j.pec.2004.09.008
16. Seeger A-M, Pfeiffer JJ, Heinzl A. When do we need a human? anthropomorphic design and trustworthiness of conversational agents. In: Proceedings of the 16th Annual Pre-ICIS Workshop on HCI Research in MIS, 2017 December 10, Seoul, Korea. (2017). <http://aisel.aisnet.org/sighci2017/15> (accessed February 1, 2021).
17. Diederich S, Brendel AB, Kolbe LM. Designing anthropomorphic enterprise conversational agents. *Bus Inf Syst Eng.* (2020) 62:193–209. doi: 10.1007/s12599-020-00639-y
18. Go E, Sundar SS. Humanizing chatbots: the effects of visual, identity and conversational cues on humanness perceptions. *Comput Human Behav.* (2019) 97:304–16. doi: 10.1016/j.chb.2019.01.020
19. Sundar SS, Oeldorf-Hirsch A, Garga AK. “A cognitive-heuristics approach to understanding presence in virtual environments,” in: Anna S, Luciano G, editors. *Proceedings of the 11th Annual International Workshop on Presence, 16-18 October, Padova, Italy, 2008.* (2008).p. 16–8.
20. Luxton DD, Sirotnin A. “Intelligent conversational agents in global health,” In: Okpaku S, editor. *Innovations in Global Mental Health.* New York, NY: Springer International Publishing. (2020). p. 1–12. doi: 10.1007/978-3-319-70134-9\_11-1
21. Yin L, Bickmore T, Cortés DE. “The impact of linguistic and cultural congruity on persuasion by conversational agents,” *Lect Notes Comput Sci.* (2010) 6356 LNAI:343–9. doi: 10.1007/978-3-642-15892-6\_36
22. Kowatsch T, Schachner T, Harperink S, Barata F, Dittler U, Xiao G, et al. conversational agents as mediating social actors in chronic disease management involving healthcare professionals, patients, and family members: intervention design and results of a multi-site, single-arm feasibility study (Preprint). *J Med Internet Res.* (2020) 23:25060. doi: 10.2196/preprints.25060
23. Bonnevie E, Lloyd TD, Rosenberg SD, Williams K, Goldberg J, Smyser J. Layla’s got you: developing a tailored contraception chatbot for black and hispanic young women. *Health Educ J.* (2020) 80:413–24. doi: 10.1177/0017896920981122
24. Rheu M, Shin JY, Peng W, Huh-Yoo J. Systematic review: trust-building factors and implications for conversational agent design. *Int J Hum Comput Interact.* (2021) 37:81–96. doi: 10.1080/10447318.2020.1807710
25. Holtgraves TM, Ross SJ, Weywadt CR, Han TL. Perceiving artificial social agents. *Comput Hum Behav.* (2007) 23:2163–74. doi: 10.1016/j.chb.2006.02.017
26. Skjuve M, Folstad A, Fostervold KI, Brandtzaeg PB. My chatbot companion - a study of human-chatbot relationships. *Int J Hum Comput Stud.* (2021) 149:102601. doi: 10.1016/j.ijhcs.2021.102601
27. Bickmore T, Cassell J. Small talk and conversational storytelling in embodied conversational interface agents. *Proc AAAI Fall Symp Narrat Intell.* (1999) :87–92.
28. Fussell SR, Kiesler S, Setlock LD, Yew V. “How people anthropomorphize robots,” In: *HRI 2008 - Proceedings of the 3rd ACM/IEEE International Conference on Human-Robot Interaction: Living with Robots.* New York, NY: ACM Press. (2008) p. 145–52.
29. Ashfaq M, Yun J, Yu S, Loureiro SMC. Modeling the determinants of users’ satisfaction and continuance intention of AI-powered service agents. *Telemat Inform.* (2020) 54:101473. doi: 10.1016/j.tele.2020.101473
30. Lee S, Lee N, Sah YJ. Perceiving a mind in a chatbot: effect of mind perception and social cues on co-presence, closeness, and intention to use. *Int J Hum Comput Interact.* (2020) 36:930–40. doi: 10.1080/10447318.2019.1699748
31. Araujo T. Living up to the chatbot hype: the influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Comput Hum Behav.* (2018) 85:183–9. doi: 10.1016/j.chb.2018.03.051
32. Verhagen T, van Nes J, Feldberg F, van Dolen W. Virtual customer service agents: using social presence and personalization to shape online service encounters. *J Comput Commun.* (2014) 19:529–45. doi: 10.1111/jcc4.12066
33. Bickmore TW, Picard RW. Establishing and maintaining long-term human-computer relationships. *Trans Comput Hum Interact.* (2005) 12:293–327. doi: 10.1145/1067860.1067867
34. Lee YC, Yamashita N, Huang Y, Fu W. “I hear you, i feel you: encouraging deep self-disclosure through a chatbot,” In: *Conference on Human Factors in Computing Systems—Proceedings.* New York, NY: ACM (2020). p. 1–12.
35. Holtgraves TM. *Language as social action: social psychology and language use.* Hove: Psychology Press (2002).
36. Holtgraves TM, Kashima Y. Language, meaning, and social cognition. *Personal Soc Psychol Rev.* (2008) 12:73–94. doi: 10.1177/1088868307309605
37. Wierzbicka A. “Anglo scripts against “putting pressure” on other people and their linguistic manifestations,” In: *Ethnopragnatics. De Gruyter Mouton.* (2011). p. 31–64.
38. Rendle-Short J. The address term mate in australian english: is it still a masculine term?. *Austr J Linguistics.* (2009) 29:245–68. doi: 10.1080/07268600902823110
39. Bacchini S. The routledge handbook of language and culture. *Ref Rev.* (2016) 30:29–31. doi: 10.1108/RR-06-2016-0145
40. Lenoir A-SI, Puntoni S, van Osselaer SMJ. “What shall I call thee? the impact of brand personality on consumer response to formal and informal address,” In: Cotte J, Wood S DM, editors. *NA - Advances in Consumer Research Volume 42.* (2014). p. 136–40. Available online at: <https://www.acrwebsite.org/volumes/1017126/volumes/v42/NA-42> [Accessed February 17, 2021]
41. Ryabova M. Politeness strategy in everyday communication. *Proced Soc Behav Sci.* (2015) 206:90–5. doi: 10.1016/j.sbspro.2015.10.033
42. Levshina N. A multivariate study of t/v forms in european languages based on a parallel corpus of film subtitles. *Res Lang.* (2017) 15:153–72. doi: 10.1515/rela-2017-0010
43. House J, Kádár DZ. T/V pronouns in global communication practices: the case of IKEA catalogues across linguacultures. *J Pragmat.* (2020) 161:1–15. doi: 10.1016/j.pragma.2020.03.001
44. Wierzbicka A. T”rms of address in european languages: a study in cross-linguistic semantics and pragmatics,” In: Allan K, Capone A, Kecskes I, editors. *Pragmatics and Theories of Language Use.* Cham: Springer International Publishing. (2016). p. 209–38.
45. Svennevig J. *Getting Acquainted in Conversation. Pragmatics and Beyond.* Amsterdam: John Benjamins Publishing Company. (2000).
46. Coveney A. Vouvoiement et tutoiement: Sociolinguistic reflections. *J French Lang Stud.* (2010) 20:127–50. doi: 10.1017/S0959269509990366
47. Dewaele J-M. Vous or tu? native and non-native speakers of french on a sociolinguistic tightrope. *Int Rev Appl Linguist Lang Teach.* (2004) 42:383. doi: 10.1515/iral.2004.42.4.383
48. Brown R, Gilman A. “The pronouns of power and solidarity,” In: Fishman JA, editor. *Readings in the Sociology of Language.* Berlin, Boston: De Gruyter. (1960) p. 252–75.
49. Wierzbicka A. Making sense of terms of address in European languages through the natural semantic metalanguage (NSM). *Intercult Pragmat.* (2016) 13:499–527. doi: 10.1515/ip-2016-0022
50. Clyne M, Norrby C, Warren J. “Language and human relations: styles of address in contemporary language,” In: *Language and Human Relations: Styles*

- of Address in *Contemporary Language*. Cambridge: Cambridge University Press. (2009). p. 1–183.
51. Solomon MR, Surprenant C, Czepiel JA, Gutman EG. A role theory perspective on dyadic interactions: the service encounter. *J Mark.* (1985) 49:99–111. doi: 10.1177/002224298504900110
  52. Clyne M, Kretzenbacher HL, Norrby C, Schüpbach D. Perceptions of variation and change in German and Swedish address. *J Socioling.* (2006) 10:287–319. doi: 10.1111/j.1360-6441.2006.00329.x
  53. Künzli A. Address pronouns as a problem in French–Swedish translation and translation revision. *Babel Rev Int la Trad / Int J Transl.* (2009) 55:364–80. doi: 10.1075/babel.55.4.04kun
  54. Warren J. Address pronouns in French: Variation within and outside the workplace. *Aust Rev Appl Linguist.* (2006) 29:1–17. doi: 10.2104/ara10616
  55. Schupbach D, Hajek J, Warren J, Clyne M, Kretzenbacher H-L, Norrby C. "A cross-linguistic comparison of address pronoun use in four European languages: intralingual and interlingual dimensions," In: *Selected Papers from the 2006 Annual meeting of the Australian Linguistic Society. School of English, Media and Art History, University of Queensland.* (2007) Available online at: <https://findanexpert.unimelb.edu.au/scholarlywork/291864-a-cross-linguistic-comparison-of-address-pronoun-use-in-four-european-languages--intralingual-and-interlingual-dimensions> (accessed February 17, 2021).
  56. Norrby C, Hajek J. "Chapter 15. Language Policy in Practice: What Happens When Swedish IKEA and H&M Take 'You' On?," In: *Uniformity and Diversity in Language Policy. Multilingual Matters.* (2011). P. 242–57.
  57. Kluge B, Moyna MI, Simon HJ, Warren J. *It's Not All About You: New Perspectives On Address Research [Internet].* (2019). doi: 10.1075/tar.1
  58. Williams L, van Compernelle RA. On versus tu and vous: Pronouns with indefinite reference in synchronous electronic French discourse. *Lang Sci.* (2009) 31:409–27. doi: 10.1016/j.langsci.2007.11.001
  59. Willis JR. "Language and identity," In: *Asian American X: An Intersection of 21st Century Asian American Voices.* London: Palgrave Macmillan. (2010). p. 214–21.
  60. Collier P, Callero P. Role theory and social cognition: learning to think like a recycler. *Self Identity.* (2005) 4:45–58. doi: 10.1080/13576500444000164
  61. Hsieh SH, Tseng TH. Playfulness in mobile instant messaging: examining the influence of emoticons and text messaging on social interaction. *Comput Human Behav.* (2017) 69:405–14. doi: 10.1016/j.chb.2016.12.052
  62. Hu Y, Wood JF, Smith V, Westbrook N. Friendships through im: examining the relationship between instant messaging and intimacy. *J Comput Commun.* (2004) 10:231. doi: 10.1111/j.1083-6101.2004.tb00231.x
  63. Wierzbicka A. A whole cloud of culture condensed into a drop of semantics. *Int J Lang Cult.* (2015) 2:1–37. doi: 10.1075/ijoc.2.1.01wie
  64. Eysenbach G. Improving the quality of web surveys: the checklist for reporting results of internet e-surveys (CHERRIES). *J Med Internet Res.* (2004) 6:e34. doi: 10.2196/jmir.6.3.e34
  65. Gefen D, Straub D. Managing user trust in B2C e-services. *e-Service J.* (2003) 2:7–24. doi: 10.2979/esj.2003.2.2.7
  66. Wixom BH, Todd PA. A theoretical integration of user satisfaction and technology acceptance. *Inf Syst Res.* (2005) 16:85–102. doi: 10.1287/isre.1050.0042
  67. Grohmann B. Gender dimensions of brand personality. *J Mark Res.* (2009) 46:105–9. doi: 10.1509/jmkr.46.1.105
  68. Mende M, van Doorn J. Coproduction of transformative services as a pathway to improved consumer well-being. *J Serv Res.* (2015) 18:351–68. doi: 10.1177/1094670514559001
  69. Büttgen M, Schumann JH, Ates Z. Service locus of control and customer coproduction. *J Serv Res.* (2012) 15:166–81. doi: 10.1177/1094670511435564
  70. Kehr F, Kowatsch T, Wentzel D, Fleisch E. Blissfully ignorant: the effects of general privacy concerns, general institutional trust, and affect in the privacy calculus. *Inf Syst J.* (2015) 25:607–35. doi: 10.1111/isj.12062
  71. Reichheld FF. The one number you need to grow. *Harvard Bus Rev.* (2003) 81:46–55.
  72. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *Mis Q.* (1989) 13:319–40. doi: 10.2307/249008
  73. Liao QV, Hussain MMU, Chandar P, Davis M, Khazaen Y, Crasso MP, et al. "All work and no play? Conversations with a question-and-answer chatbot in the wild," In: *Proceedings of the Conference on Human Factors in Computing Systems.* Association for Computing Machinery. (2018).
  74. Oh C, Song J, Choi J, Kim S, Lee S, Suh B. "I lead, you help but only with enough details: Understanding the user experience of co-creation with artificial intelligence," In: *Proceedings of the Conference on Human Factors in Computing Systems.* New York, NY: ACM. (2018). p. 1–13.
  75. Shamekhi A, Liao QV, Wang D, Bellamy RKE, Erickson T. "Face value? exploring the effects of embodiment for a group facilitation agent," In: *Proceedings of the Conference on Human Factors in Computing Systems.* New York, NY: ACM. (2018) p. 1–13. doi: 10.1145/3173574.3173965
  76. Bland JM, Altman DG. Statistics notes: cronbach's alpha. *BMJ.* (1997) 314:572. doi: 10.1136/bmj.314.7080.572
  77. Altman DG, Royston P. Statistics notes: the cost of dichotomising continuous variables. *BMJ Br Med J.* (2006) 332:1080. doi: 10.1136/bmj.332.7549.1080
  78. Sarma KVS, Vardhan RV. *Multivariate Statistics Made Simple: A Practical Approach.* London: CRC Press. (2018)
  79. Yzerbyt VY, Muller D, Judd CM. Adjusting researchers' approach to adjustment: on the use of covariates when testing interactions. *J Exp Soc Psychol.* (2004) 40:424–31. doi: 10.1016/j.jesp.2003.10.001
  80. Field A, Miles J, Field Z. *Discovering Statistics Using R.* London: Sage London. (2012).
  81. Tabachnick BG, Fidell LS, Ullman JB. *Using Multivariate Statistics.* Boston, MA: Pearson. (2007).
  82. Stevens JP. *Applied Multivariate Statistics for the Social Sciences.* London: Routledge. (2012).
  83. Cohen J. *Statistical Power Analysis for the Behavioral Sciences.* London: Academic press. (1988).
  84. Ring L, Utami D, Bickmore T. The right agent for the job? *Lect Notes Comput Sci.* (2014) 49:374–84. doi: 10.1007/978-3-319-09767-1\_49
  85. Branaghan RJ, Hildebrand EA. Brand personality, self-congruity, and preference: a knowledge structures approach. *J Consum Behav.* (2011) 10:304–12. doi: 10.1002/cb.365
  86. Ami D, Arahamian F, Luchini S. Stated preferences and decision-making: three applications to health. *Rev économique.* (2017) 68:327. doi: 10.3917/reco.683.0327
  87. Lambooi MS, Harmsen IA, Veldwijk J, de Melker H, Mollema L, van Weert YW, et al. Consistency between stated and revealed preferences: a discrete choice experiment and a behavioural experiment on vaccination behaviour compared. *BMC Med Res Methodol.* (2015) 15:19. doi: 10.1186/s12874-015-0010-5
  88. Gosselin K, Norris JL, Ho MJ. Beyond homogenization discourse: reconsidering the cultural consequences of globalized medical education. *Med Teach.* (2016) 38:691–9. doi: 10.3109/0142159X.2015.1105941
  89. Mangham LJ, Hanson K. Scaling up in international health: what are the key issues? *Health Policy Plan.* (2010) 25:85–96. doi: 10.1093/heapol/czp066
  90. Kowatsch T, Nißen M, Rügger D, Stieger M, Flückiger C, Allemann M, et al. *The impact of interpersonal closeness cues in text-based healthcare chatbots on attachment bond and the desire to continue interacting: an experimental design.* (2018).
  91. Reichheld FF, Covey S, Mekonnen A. The ultimate question: driving good profits and true growth. *J Target Measure Anal Market.* (2006) 14:369–70. doi: 10.1057/palgrave.jt.5740195
  92. Hauser-Ulrich S, Künzli H, Meier-Peterhans D, Kowatsch T. A smartphone-based health care chatbot to promote self-management of chronic pain (selma): pilot randomized controlled trial. *JMIR mHealth uHealth.* (2020) 8:1–23. doi: 10.2196/15806
  93. Schwarzer R. Modeling health behavior change: how to predict and modify the adoption and maintenance of health behaviors. *Appl Psychol.* (2008) 57:1–29. doi: 10.1111/j.1464-0597.2007.00325.x
  94. Sheeran P, Webb TL. The intention-behavior gap. *Soc Personal Psychol Compass.* (2016) 10:503–18. doi: 10.1111/spc3.12265
  95. Barello S, Triberti S, Graffigna G, Libreri C, Serino S, Hibbard J, et al. EHealth for patient engagement: a systematic review. *Front Psychol.* (2016) 6:2013. doi: 10.3389/fpsyg.2015.02013
  96. Kuzelewska E. Language policy in Switzerland. *Stud Logic Gramm Rhetor.* (2016) 45:125–40. doi: 10.1515/slgr-2016-0020

97. Watts RJ. Language, dialect and national identity in Switzerland. *Multilingua*. (1988) 7:313–4. doi: 10.1515/mult.1988.7.3.313
98. Rietz T, Benke I, Maedche A. "The impact of anthropomorphic and functional chatbot design features in enterprise collaboration systems on user acceptance." In: *Wirtschaftsinformatik 2019 Proc.* (2019) Available online at: <https://aisel.aisnet.org/wi2019/track13/papers/7> (accessed July 30, 2021).

**Conflict of Interest:** JO, MN, and FvW are affiliated with the Center for Digital Health Interventions ([www.c4dhi.org](http://www.c4dhi.org)), a joint initiative of the Department of Management, Technology, and Economics at ETH Zurich and the Institute of Technology Management at the University of St. Gallen, which is funded in part by the Swiss health insurance CSS Versicherungen. CSS provided funding for recruiting participants but was neither involved in any aspect of the study design, data analysis nor manuscript preparation.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Ollier, Nißen and von Wangenheim. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.