

Is voice really persuasive? The influence of modality in virtual assistant interactions and two alternative explanations

Carolin Ischen, Theo B. Araujo, Hilde A.M. Voorveld, Guda Van Noort and Edith G. Smit

*Amsterdam School of Communication Research ASCoR,
Universiteit van Amsterdam, Amsterdam, The Netherlands*

Abstract

Purpose – Virtual assistants are increasingly used for persuasive purposes, employing the different modalities of voice and text (or a combination of the two). In this study, the authors compare the persuasiveness of voice and text-based virtual assistants. The authors argue for perceived human-likeness and cognitive load as underlying mechanisms that can explain why voice- and text-based assistants differ in their persuasive potential by suppressing the activation of consumers' persuasion knowledge.

Design/methodology/approach – A pre-registered online-experiment ($n = 450$) implemented a text-based and two voice-based (with and without interaction history displayed in text) virtual assistants.

Findings – Findings show that, contrary to expectations, a text-based assistant is perceived as more human-like compared to a voice-based assistant (regardless of whether the interaction history is displayed), which in turn positively influences brand attitudes and purchase intention. The authors also find that voice as a communication modality can increase persuasion knowledge by being cognitively more demanding in comparison to text.

Practical implications – Simply using voice as a presumably human cue might not suffice to give virtual assistants a human-like appeal. For the development of virtual assistants, it might be beneficial to actively engage consumers to increase awareness of persuasion.

Originality/value – The current study adds to the emergent research stream considering virtual assistants in explicitly exploring modality differences between voice and text (and a combination of the two) and provides insights into the effects of persuasion coming from virtual assistants.

Keywords Virtual assistant, Modality, Persuasion knowledge, Human-likeness, Cognitive load

Paper type Research paper

The use of intelligent virtual assistants, virtual characters that emulate human interaction, is rapidly increasing (Gray, 2016). Virtual assistants have an enormous potential for business as they can pay 24/7 attention to consumer questions and requests and provide product- or service-related recommendations in a natural manner (Accenture, 2018). Investments in text-based technologies (e.g. chatbots) have been growing over the last years, and the global market is predicted to grow even further (Grand View Research Inc, 2021; Research and Markets, 2021). Moreover, businesses lately also pay increasing attention to voice-based technologies (such as those implemented on Amazon Alexa or Google Assistant). These are



not only an interactive way for consumers to obtain daily information, e.g. about weather, traffic or news, but can also enable new ways of providing persuasive (e.g. brand-related) messages (Duggal, 2020; Tillman and O'Boyle, 2019).

Research has started to widely investigate the social cues that drive perceptions of virtual assistants and their influence on different social, relational and persuasive outcomes (e.g. Cheng *et al.*, 2021; Lee *et al.*, 2021; Rhee and Choi, 2020). By doing so, existing research often focuses on social cues that are specific to one modality, e.g. visual cues in text-based assistants (e.g. de Visser *et al.*, 2016; Forlizzi *et al.*, 2007; Nowak and Biocca, 2003), or auditory cues such as emotional tone, vocal pitch, etc. in voice-based assistants (e.g. Ding *et al.*, 2014; Louwerse *et al.*, 2005, for an overview of social cues see Feine *et al.*, 2019).

Despite research comparing different cues within one modality, little attention has been devoted to a comparison between different modalities in virtual assistants, especially regarding persuasion. This is surprising given that voice-based assistants might be more persuasive in comparison to text-based assistants, because they verbally deliver product or service recommendations in a natural and playful manner (Cox, 2018; Moriuchi, 2019). They possibly imbue voice as a social cue that provides a strategy to give the interaction a more “human touch” (Besik, 2019; Sivaramakrishnan *et al.*, 2007).

Hence, given their popularity and possible persuasiveness, it is important for theorists, researchers, marketers, regulators and policymakers to know how consumers respond to virtual assistants implementing voice (vs text, or a combination of the two) for disseminating persuasive messages. The current research aims to make an important contribution to the field of marketing communication by explicitly comparing the different virtual assistant modalities of voice, text or a combination, in their persuasive potential. The overall research question is: *Does using voice instead of, or in addition to text make a virtual assistant more persuasive?*

From a theoretical standpoint, the current research examines voice as a factor that challenges consumers' ability to recognize and cope with persuasive attempts, especially when embedded in a conversation. It takes two underlying mechanisms into account. Firstly, the concept of human-likeness is one of the most important mechanisms explored in human-machine communication research (Rapp *et al.*, 2021), as humans tend to imbue nonhuman agents with human-like characteristics (Epley *et al.*, 2007). It can be argued that voice in itself (vs text) might be the element that gives virtual assistants its human-like appeal (Cho *et al.*, 2019). This possibly increases its social character and in turn its persuasiveness. Secondly, it has been debated whether speaking is an easier and more natural way to communicate with a virtual assistant than using text-based input (Cox, 2018). However, research has also shown that voice can be less efficient and increase mental demand in comparison to text (Le Bigot *et al.*, 2004). Furthermore, if interacting via voice is more demanding for consumers, it might leave less (mental) room to identify persuasive attempts. Hence, cognitive load plays an important role when examining persuasion (Berry *et al.*, 2005; Van Zant and Berger, 2019). Therefore, this research includes perceived human-likeness and cognitive load as two alternative mechanisms that can explain the persuasiveness of virtual assistants.

This research provides important theoretical, practical and societal contributions. It makes a theoretical contribution by adding to knowledge on the overall effectiveness of different virtual assistant modalities. As businesses increasingly use voice-based virtual assistants to disseminate persuasive messages, it is important to understand if voice-based technologies (in comparison to text) can influence whether consumers are aware of a virtual assistant's persuasive potential. This knowledge makes a practical contribution by helping businesses in increasing their profitability when using virtual assistants as a marketing tool. At the same time, it (societally) contributes to consumer empowerment by exploring whether consumers can discern commercial and non-commercial messages when using virtual assistants.

Virtual assistants and perceived human-likeness

Human-likeness is a key concept in the field of human-machine communication (Guzman and Lewis, 2020). Several studies indicate that humans ascribe human characteristics to nonhuman agents such as virtual assistants (Epley et al., 2007; for a literature review see Rapp et al., 2021). However, very little is known about the extent to which the way in which we communicate with a virtual assistant, i.e. the modality (voice-based and/or text-based), influences our perceptions of human-likeness. To the best of our knowledge, only one study so far has examined differences in perceived human-likeness between voice and text modality of the same virtual assistant source. Cho et al. (2019) found that voice (vs text) is perceived as more human-like, subsequently leading to more positive attitudes toward a virtual assistant for utilitarian tasks. As virtual assistants become increasingly prevalent in the marketing field to influence consumers, our research aims to add to this explicit modality comparison by applying it to a persuasion context.

We conceptualize perceived human-likeness as a combination of anthropomorphism or humanness of a technology (adapted from the humanness index) and social presence (Cho et al., 2019). Anthropomorphism is the attribution of human qualities to the virtual assistant (such as friendliness or lifelikeness, Kim and Sundar, 2012), while social presence taps into the perception of communicating with a social interaction partner (Lee, 2004). Even though there are conceptual differences in the literature, they are closely related as they both relate to the sociability or the human touch of a (virtual) entity.

We draw on the Computers are Social Actors paradigm (Reeves and Nass, 1996), stating that humans apply social rules to interactions with technology similarly to interactions with other human beings. Following this paradigm, using social cues in technology interactions enhances social responses. Given that the interaction with voice is an inherently human characteristic (Pinker and Bloom, 1990), a virtual assistant emulating human voice is expected to be perceived as more human-like in comparison to communicating via text only (see Figure 1). This assumption is supported by Schroeder and Epley (2016), who found in a series of experiments that paralinguistic cues in speech (e.g. pace, intonation) influence people’s perceptions of human-likeness in comparison to text-based interaction. To extend these previous findings, we firstly examine whether voice itself (in comparison to text) enhances perceptions of human-likeness. We propose the following:

H1. A voice-based virtual assistant is perceived as higher in human-likeness than a text-based virtual assistant.

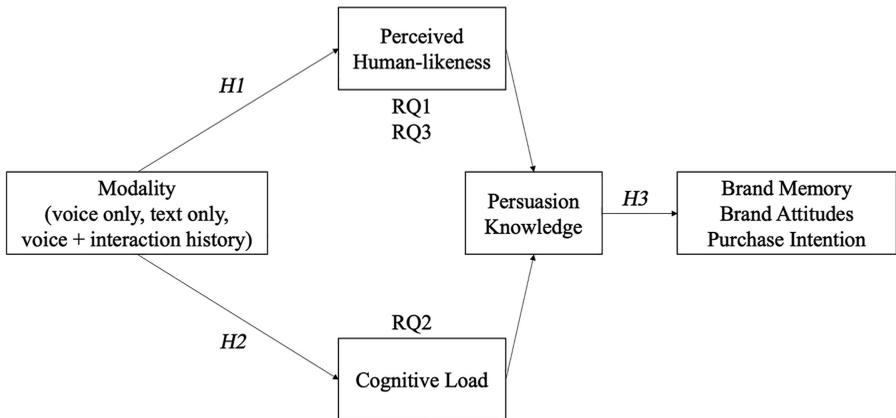


Figure 1. Conceptual mediation model for persuasion knowledge, and advertising effectiveness in virtual assistants

Virtual assistants and persuasion knowledge

A large body of literature has been devoted to understanding how people recognize, process and respond to persuasion techniques that are subtler than traditional advertising, such as sponsored content, brand placement or advergames (Boerman *et al.*, 2012; Tutaj and van Reijmersdal, 2012). Building on the persuasion knowledge model of Friestad and Wright (1994), one central claim of this line of research is that these formats include hidden persuasive attempts that are less identifiable than traditional advertising. Studying these more subtle forms of advertising is important considering consumer empowerment, as consumers might be more prone to persuasive attempts that can potentially be misleading.

Virtual assistants can embed a persuasive attempt in their messages, e.g. by advertising brands in their communication. The persuasion knowledge model is therefore a useful anchor point to develop a research model for explaining the persuasive effects of advertising delivered by a virtual assistant. The concept of persuasion knowledge describes a range of competences that are related to the understanding of advertising in general, and the persuasive intent of advertising more specifically (Tutaj and van Reijmersdal, 2012). Persuasion knowledge can be developed through experience with the persuasion technique and through socialization (Friestad and Wright, 1994). While consumers might already have developed some persuasion knowledge with regard to other persuasive techniques (e.g. sponsored content, celebrity endorsement; Boerman *et al.*, 2017), it can be assumed that the persuasion knowledge about virtual assistants as a new technique is less refined.

While originally conceptualized as a dispositional variable (Friestad and Wright, 1994), several scholars have shown interest in situations that can activate higher or lower levels of persuasion knowledge following a specific persuasion tactic (Boerman *et al.*, 2012; Campbell, 1995; van Noort *et al.*, 2012; Tutaj and van Reijmersdal, 2012). Within the broad range of conceptualizations in literature and empirical studies (for an overview, see Boerman *et al.*, 2018; Ham *et al.*, 2015), the current study specifically applies consumers' understanding of a persuasive or selling intent in an online advertising context to the context of virtual assistants. Hence, persuasion knowledge is treated as a situational variable (van Noort *et al.*, 2012; Tutaj and van Reijmersdal, 2012).

Perceived human-likeness as the primary explanation for reduced persuasion knowledge

This research studies perceived human-likeness as the primary underlying mechanism explaining the effect of modality on reduced persuasion knowledge. Examining the relationship of perceived human-likeness and persuasiveness is crucial, as virtual assistants are characterized by their social or human-like cues that can potentially make them more influential. Research on paralinguistic cues builds on the idea that voice, in comparison to text, provides more opportunities for the illustration of human traits, states and feelings in the communication (Van Zant and Berger, 2019). This line of research has found voice to influence persuasiveness via positive perceptions of these human characteristics (i.e. having a confident appearance; Van Zant and Berger, 2019).

Moreover, Campbell and Kirmani (2000) identified the accessibility of ulterior motives as a key factor responsible for whether persuasion knowledge is used. In other words, if consumers infer that the underlying motive for the conversation with the virtual assistant is persuasive (rather than social) in nature, they are more likely to activate persuasion knowledge. This applies especially in a situation in which a persuasive attempt is embedded in a conversation. We argue that when the interaction via voice is perceived as more human-like, virtual assistant users apply social responses toward it (following the CASA paradigm; Reeves and Nass, 1996), including greater social attractiveness (Lee, 2010).

Consumers might thus be more likely to infer social motives for the interaction with the virtual assistant such as relationship building or helping, rather than a persuasive motive. Hence, the perception of human-likeness might lead to lower persuasion knowledge. However, since empirical evidence is lacking to formulate a supported hypothesis, we propose the following research question:

- RQ1.* Does interacting with a voice-based virtual assistant lead to lower persuasion knowledge in comparison to interacting with a text-based virtual assistant mediated by higher perceptions of the virtual assistant's human-likeness?

Cognitive load as an alternative explanation

As an alternative explanation, this study includes cognitive load. Cognitive load has been shown to play an important role when examining persuasion (e.g. [Berry et al., 2005](#)). The concept represents the mental burden that a particular task imposes on a user's cognitive system ([Paas and Van Merriënboer, 1994](#)). Text is self-paced (and thus more controllable) and allows to go back and forth in the interaction (for text-effects in multimedia learning, see [Schmidt-Weigand et al., 2010](#)). Based on cognitive load theory (see [Leahy and Sweller, 2011](#)), it is likely that interacting via text imposes a lower cognitive load on the user compared to voice that is more demanding. This notion has been supported by a study on virtual assistants by [Berry et al. \(2005\)](#), finding text as being easier understood than voice. Further, research on modality differences (in for example the word-of-mouth context) showed that the asynchronous nature of written communication allows greater time to construct and refine what to say ([Berger and Iyengar, 2013](#)). To test this assumption, we propose the following [1]:

- H2.* Cognitive load is higher for interacting with a voice-based virtual assistant than for interacting with a text-based virtual assistant.

For persuasion knowledge to be activated and utilized, information has to be retrieved from memory ([Campbell and Kirmani, 2000](#); [Hossain and Saini, 2014](#)). This makes cognitive processing the second primary antecedent ([Kirmani and Campbell, 2008](#)), for both, recognizing the persuasive attempt and responding to it ([Campbell, 1995](#); [Hossain and Saini, 2014](#)). Research showed that cognitively busy people are less likely to activate and use their persuasion knowledge in a given situation ([Campbell and Kirmani, 2000](#)). Applied to the virtual assistant interaction, this means that the higher the cognitive load imposed on a consumer by communicating with the assistant itself, the less cognitive processing of the actual (content of) interaction occurs. If this is the case, the consumer is then also less likely to activate their persuasion knowledge. The effect of voice on persuasion knowledge might then be attributed to differences in cognitive load, complementing perceived human-likeness of the voice-based virtual assistant. To test this assumption, we propose the following research question including both, perceived human-likeness, and cognitive load:

- RQ2.* Does interacting with a voice-based virtual assistant lead to lower persuasion knowledge in comparison to interacting with a text-based virtual assistant, mediated by both, higher perceptions of the virtual assistant's human-likeness and higher cognitive load?

Interaction history reducing cognitive load

Virtual assistants can implement the display of interaction history in addition to using voice, as for example done in smart displays ([Seifert, 2020](#)). In other words, everything the voice-based virtual assistant delivers verbally is translated into text and shown on the screen.

Hence, cognitive load can be lifted in both, the communication with the text-based, and the communication with the voice-based virtual assistant. To make sure that the effects proposed can be attributed to the communication modality, we need to examine as well whether interacting with voice, but seeing the interaction history, results in the same effects as voice. This also allows us to study a wider array of virtual assistant modalities that are implemented in practice (e.g. smart displays).

Van Zant and Berger (2019) proposed that the human-likeness of paralinguistic cues might increase persuasion, but that these effects might disappear in the presence of linguistic cues that facilitate the detection (e.g. displaying the interaction in text on the screen). In other words, when cognitive load is lower, consumers have more cognitive resources left to detect persuasive intents coming from the virtual assistant. Further supported by classical modality studies (e.g. Pfau, 1990), text-based communication triggers people to be more focused on content-characteristics and less distracted by source-characteristics. A further explanation for cognitive load being lifted when voice-based communication is accompanied by text lies in the dual coding theory (Paivio, 1986). Dual coding theory states that verbal and visual information are coded differently, and additive effects exist for both types of codes. For the current study this means that the combination of verbal (i.e. voice) and visual (i.e. text) is easier to comprehend than text alone. Hence, to test whether the mediating effect of perceived human-likeness still holds when cognitive load is lifted in the voice-based virtual assistant through the display of interaction history, we propose the following:

- RQ3. Does interacting with a voice-based virtual assistant supported by a text-based interaction history lead to lower persuasion knowledge in comparison to interacting with a text-based virtual assistant, mediated by higher perceptions of the virtual assistant's human-likeness?

Persuasion knowledge and advertising effectiveness

Virtual assistants are used in marketing to disseminate persuasive messages and ultimately, to increase advertising effectiveness. Hence it is imperative to understand how persuasive knowledge and advertising effectiveness are related in the context of virtual assistants implementing different modalities. Persuasion knowledge may exert different effects on different types of responses. To examine the subsequent effects of persuasion knowledge on advertising effectiveness, we include three brand-related outcomes in our research model: affective (brand attitudes), cognitive (brand memory) and behavioral (purchase intention) outcomes.

We suggest that increased persuasion knowledge positively influences cognitive outcomes such as brand memory. To retrieve and utilize persuasion knowledge, people need to elaborately process the communicated message (Buijzen *et al.*, 2010). Higher persuasion knowledge then in turn also increases the likelihood of remembering the communicated brands, since cognitive processes are activated. Research on sponsorship disclosure showed a positive relationship between understanding that a message is persuasive and recognition and recall of an advertisement (e.g. Boerman and van Reijmersdal, 2020).

However, this line of research also showed that understanding a persuasive or selling intent negatively influences more affective processes (Boerman *et al.*, 2017; Boerman and van Reijmersdal, 2020) and behavioral intentions (Choi *et al.*, 2018). Thus, we propose persuasion knowledge to negatively influence brand attitudes and purchase intention as indicators for advertising effectiveness. As the persuasion knowledge model (Friestad and Wright, 1994) and theories on reactance (for an overview, see Fransen *et al.*, 2015) suggest, people use their

persuasion knowledge to cope with a persuasive attempt. Hence, persuasion knowledge enhances the critical assessment of advertising that might be perceived as a threat to peoples' individual freedom (Brehm, 1966). Even though not all previous studies found effects of persuasion knowledge on brand-related outcomes (e.g. Boerman *et al.*, 2012; Van Reijmersdal *et al.*, 2012), several studies indicated that in case a persuasive attempt is detected as such, people are more likely to critically assess the attempt (Obermiller *et al.*, 2005), negatively evaluate it (Tutaj and van Reijmersdal, 2012) and develop less positive attitudes and behavioral intentions (Boerman *et al.*, 2017; van Reijmersdal *et al.*, 2016). Hence, we propose that an increase in persuasion knowledge negatively influences affective (attitudes) and behavioral (purchase intention) brand-related outcomes. In sum, this leads to the following hypothesis:

- H3. Persuasion knowledge is positively related to (a) brand memory and negatively related to (b) brand attitudes and (c) purchase intention.

Methods

Design and sample

We implemented an experimental between-subjects design with three conditions, (1) virtual assistant communicating via voice only (voice condition; voice as input and output modality), (2) virtual assistant communicating via voice, but displaying the interaction history (voice + IH condition; voice as input modality, voice accompanied by text as output modality), and (3) virtual assistant communicating via text (text condition; text as input and output modality). The study was conducted in Dutch and was pre-registered on the open science framework (OSF) [2].

An *a priori* G*Power analysis for a between-subject one-way analysis of variance (ANOVA) with three groups informed that the required sample size was 300 for an effect size of partial eta square = 0.06 ($f = 0.25$) with 98% power. This calculation was informed by the study of Cho *et al.* (2019) estimating a similar effect. However, research on virtual assistants is an emerging area of research with a very limited number of previous studies that help us estimating effect sizes. Additionally, we wanted to account for possible technical difficulties with the setup. Hence, we were striving for a larger sample size of at least 450.

Participants were recruited through an ISO-certified research company in the Netherlands, using initial quotas for age, gender and region, and using a continuous recruit stream in June and July 2020 to reach the desired sample size. Before filling in the survey, participants had to give informed consent and make sure they used Google Chrome as a browser. Participation was terminated in case participants failed one or both attention checks ($n = 305$), failed to received audio in the voice conditions ($n = 37$), or did not interact with the virtual assistant, meaning that they spent less than 15 seconds on the interactions, or could not recall the recommendation given by the assistant ($n = 398$). After excluding one participant who indicated to be under 18 years old and one multivariate outlier [3], the final sample was 450. Participants in the final sample were between 18 and 78 years old ($M = 46.15$, $SD = 15.84$), 229 were male (50.89%; 220 female, 1 non-binary). In terms of education, 54.22% indicated to have a high educational level (36.0% middle, 9.78% low). A full overview of all descriptive statistics is presented in the Appendix.

Procedure

The study was approved by the university's Ethical Review Board. After giving informed consent and randomly being assigned to the conditions ($n_{\text{Voice}} = 113$, $n_{\text{Voice} + \text{IH}} = 134$,

$n_{\text{Text}} = 203$) [4], participants were instructed to interact with the virtual assistant to obtain a recommendation for a dinner recipe. We chose this task because virtual assistants are often used for cooking-related questions and it allows for the embedding of branded product-related recommendations (Rabideau, 2018). Participants were guided through the interaction by the virtual assistant, including the request to choose one out of three pasta dishes (beef, chicken or vegetarian to account for differences in taste). After choosing a dish, the virtual assistant gave the ingredients for four portions including eleven ingredients each. The recommendation contained five branded ingredients. A full transcript of the interaction is provided on the OSF [5].

Stimuli

Three versions of a virtual assistant were designed for this study using and extending the conversational agent research toolkit (Araujo, 2020). In the voice only condition, participants were exposed to a microphone icon they had to click on to start talking to the assistant. The assistant responded via voice, providing a recommendation for a recipe. The voice used was the voice “Xander” – a synthetic younger male voice – available on Google Chrome (an example of the voice is provided on the OSF). The voice plus interaction history condition resembled the voice only condition in its graphical interface. In addition to the verbal input and output, the interaction was translated into text that was displayed in a chat-interface. In the text only condition, participants interacted with the assistant via a chat-interface. Examples of the stimuli are presented in Figure 2.

Measurements

All items were measured with a 7-point Likert-scale ranging from 1 = strongly disagree to 7 = strongly agree, unless stated otherwise. All measurements were translated from English to Dutch for the experiment.

Perceived human-likeness. *Perceived human-likeness* was assessed with a combination of mindful anthropomorphism, measured with four items on a 7-point semantic differential scale including “I perceived the virtual assistant as machine-like/human-like” (Bartneck *et al.*, 2009; Ho and MacDorman, 2010) and social presence, measured with nine items including “While I was communicating with the virtual assistant, I felt as if it was an intelligent being” (Gefen and Straub, 2004; Lee and Nass, 2003; Cronbach’s alpha = 0.96, $M = 3.94$, $SD = 1.46$).

Cognitive load. To assess *cognitive load*, we measured the amount of mental effort, referring to the cognitive capacity allocated to the communication (Paas *et al.*, 2003). We used one item “How much mental effort have you invested in the interaction with the virtual assistant?” on a 7-point scale from 1 = very low to 7 = very high (Paas, 1992; $M = 3.30$, $SD = 1.47$) [6].

Persuasion knowledge. *Persuasion knowledge*, conceptualized as the understanding of persuasive and selling intent, was measured with seven items (Tutaj and van Reijmersdal, 2012). Two items were used to measure selling intent: “The aim of this virtual assistant is to sell products” and “The aim of this virtual assistant is to stimulate the sales of products”. Two items were used to measure persuasive intent: “The aim of this virtual assistant is to influence my opinion” and “The aim of this virtual assistant is to make people like certain products”. Three items were used as filler items referring to an informational attempt (helping): “The aim of this virtual assistant is to help me choose a recipe”, “The aim of this virtual assistant is to give information about recipes” and “The aim of this virtual assistant is to let people know more about recipes”. To assess persuasion knowledge, we used the four items measuring a persuasive or selling intent, which formed a reliable scale (Cronbach’s alpha = 0.83, $M = 4.48$, $SD = 1.29$).

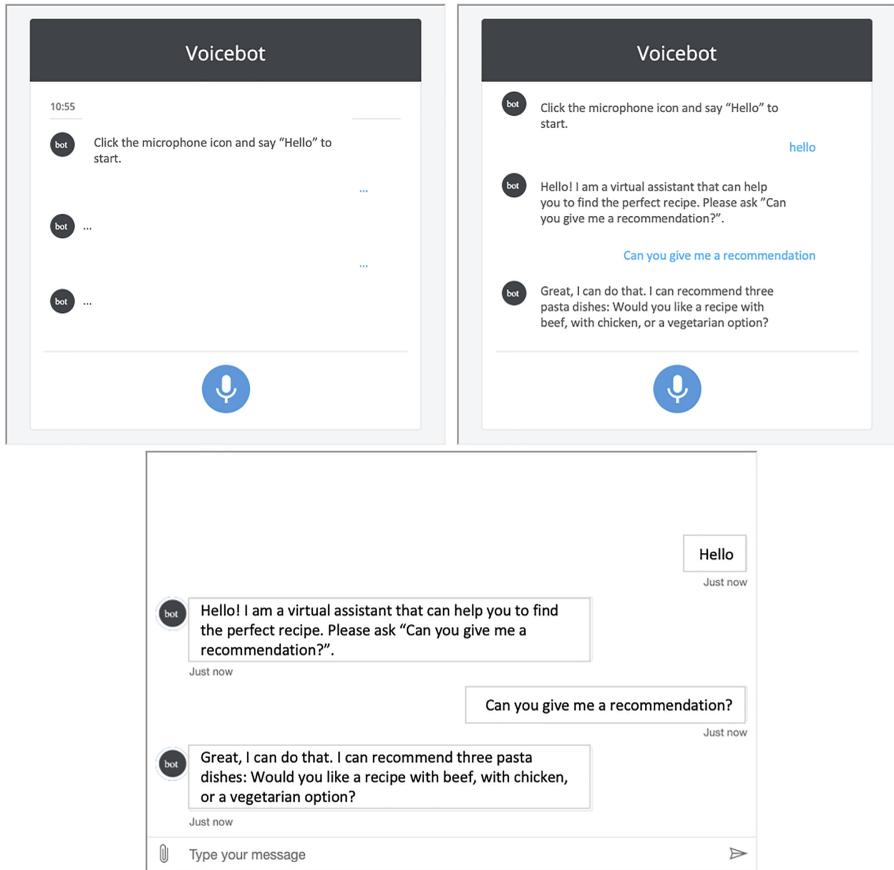


Figure 2.
Stimulus material
(voice only;
voice + interaction
history; text only),
translated to English
from Dutch

Advertising effectiveness. *Brand memory* was assessed with a recall and a recognition task. Firstly, participants were asked to write a shopping list with all items they remembered from the interaction. In the second, guided task, participants were asked to tick all items they would put on the shopping list out of a list with several options. For both, the recall, and the recognition task, we counted the total number of correctly identified brands/branded ingredients (scale from 0–5; $M_{\text{recall}} = 0.15$, $SD_{\text{recall}} = 0.53$; $M_{\text{recognition}} = 2.26$, $SD_{\text{recognition}} = 1.59$). Since most participants were unable to freely recall the brands, we exclude this variable in the subsequent analysis and only use recognition as an indicator for brand memory. *Brand attitudes* were measured with four items on a 7-point semantic-differential scale including “I think the recommended brands are negative-positive/uninteresting-interesting/unattractive-attractive/bad-good” (van Reijmersdal *et al.*, 2016; Cronbach’s alpha = 0.94, $M = 4.98$, $SD = 1.29$). *Purchase intention* was measured with four items including “I would purchase the recommended brands” (Dabholkar and Sheng, 2012; Cronbach’s alpha = 0.85, $M = 4.23$, $SD = 1.34$).

Control variables. We measured *age*, *gender* and *education*. We also measured *familiarity with the featured brands*, i.e. number of known brands (scale from 0–8, $M = 4.38$, $SD = 1.48$); and *familiarity with virtual assistants* (Zhou *et al.*, 2010; Cronbach’s alpha = 0.82, $M = 4.16$,

$SD = 1.53$). We also controlled for *preference for voice, or text-based communication* (Pastore, 2014; $M = 4.95$, $SD = 2.08$) and *preference for the recommended dish* ($M = 5.47$, $SD = 1.16$).

Is voice really persuasive?

Technical pretest

We conducted a technical pretest with 13 Master's students of Communication Science in a classroom setting, of which 10 completed an online questionnaire assessing common variables of the Technology Acceptance Model (Davis *et al.*, 1989) beforehand. Based on the feedback, we adapted the wording of the interaction dialog with the virtual assistant.

411

Results

Randomization check

To see whether the random assignment to the different conditions was successful, we conducted chi-square difference tests with *gender* and *education* as the dependent variable and the three conditions as the independent variable. The proportions for *gender* ($X^2(4) = 7.46$, $p = 0.114$) and *education* ($X^2(10) = 8.65$, $p = 0.566$) did not differ by condition.

Furthermore, we conducted one-way ANOVAs with all other control variables as the dependent variable respectively and the three conditions as the between-subjects factor. *Age* ($F(2, 447) = 0.29$, $p = 0.750$), and *liking of the recommended dish* ($F(2, 447) = 0.82$, $p = 0.443$) did not differ across conditions. We found significant differences across conditions for *familiarity with the featured brands* ($F(2, 447) = 6.67$, $p < 0.001$). Familiarity with the featured brands was significantly lower in the voice only condition ($M_{voice} = 4.02$, $SD_{voice} = 1.48$, $p < 0.001$) than in the voice + interaction history condition ($M_{voice + IH} = 4.70$, $SD_{voice + IH} = 1.37$), but not in comparison to the text condition ($M_{text} = 4.37$, $SD_{text} = 1.52$, $p = 0.103$). *Familiarity with virtual assistants* also differed across conditions ($F(2, 447) = 16.59$, $p < 0.001$). Familiarity with virtual assistants was significantly higher in the text condition ($M_{text} = 4.56$, $SD_{text} = 1.40$) than in the voice + interaction history condition ($M_{voice + IH} = 4.07$, $SD_{voice + IH} = 1.56$, $p = 0.009$) and the voice only condition ($M_{voice} = 3.57$, $SD_{voice} = 1.53$, $p < 0.001$). The mean difference between participants in the voice only and voice + interaction history condition was also significant ($p = 0.022$). *Preference for voice, or text-based communication* differed across conditions ($F(2, 447) = 28.97$, $p = 0.001$), participants in the text condition were significantly more in favor of text-based communication ($M_{text} = 5.69$, $SD_{text} = 1.81$) than participants in the voice + interaction history condition ($M_{voice + IH} = 4.08$, $SD_{voice + IH} = 2.06$, $p < 0.001$) and in the voice only condition ($M_{voice} = 4.95$, $SD_{voice} = 2.10$, $p < 0.001$) [7]. Since these three variables were also significantly correlated with at least one of the outcome variables, they were included as covariates in the subsequent analyses.

Hypothesis testing

To test H1, we conducted a one-way ANCOVA with perceived human-likeness as the dependent variable and the three conditions as the between-subjects factor, controlling for familiarity with the featured brands, familiarity with virtual assistants and preference for voice- or text-based communication. We find a significant effect of modality on perceived human-likeness ($F(2, 444) = 7.05$, $p < 0.001$, partial eta square = 0.03). However, a post-hoc comparison with Tukey HSD adjustments showed an unexpected pattern. Participants in the text condition ($M = 4.20$, $SD = 1.50$) perceived the virtual assistant significantly more human-like than participants in the voice + interaction history condition ($M = 3.84$, $SD = 1.44$, $p = 0.032$) and the voice only condition ($M = 3.59$, $SD = 1.32$, $p = 0.001$). H1 cannot be supported.

To test H2, we conducted a one-way ANCOVA with cognitive load as the dependent variable and the three conditions as the between-subjects factor, controlling for familiarity with the featured brands, familiarity with virtual assistants and preference for voice- or text-based communication. We find a significant effect of modality on cognitive load ($F(2, 444) = 3.43, p = 0.033$, partial eta square = 0.02). Confirming our expectations, cognitive load was highest in the voice only condition ($M_{\text{text}} = 3.11, SD_{\text{text}} = 1.47; M_{\text{voice + IH}} = 3.41, SD_{\text{voice + IH}} = 1.52, M_{\text{voice}} = 3.49, SD_{\text{voice}} = 1.40$). Note however that post-hoc comparisons indicated no significant differences across experimental conditions. In sum, H2 cannot be supported.

To test H3, we conducted a simple regression with persuasion knowledge (persuasive and selling intent) as the independent and brand memory, brand attitudes and purchase intention as the dependent variable respectively, controlling for familiarity with the featured brands, familiarity with virtual assistants and preference for voice- or text-based communication. A significant regression equation was found for brand memory ($F(4, 445) = 16.95, p < 0.001, R^2 = 0.13, b = 0.23$), brand attitudes ($F(4, 445) = 7.14, p < 0.001, R^2 = 0.06, b = -0.13$) and purchase intention ($F(4, 445) = 10.08, p < 0.001, R^2 = 0.08, b = -0.19$). H3 can be supported as persuasion knowledge is positively related to brand memory and negatively related to brand attitudes and purchase intention.

Analysis of the full research model

To answer the research questions and test the full proposed model, we used the PROCESS macro for SPSS (model 80) with bootstrapping (10,000 samples) to create confidence intervals for the indirect effects (Hayes, 2017). Since the independent variable is multi-categorical, we ran the analysis twice accounting for all different comparisons. The results of the full models are presented in Figure 3 (brand memory), Figure 4 (brand attitudes) and Figure 5 (purchase intention) and regression tables are presented in the Appendix. In line with the results

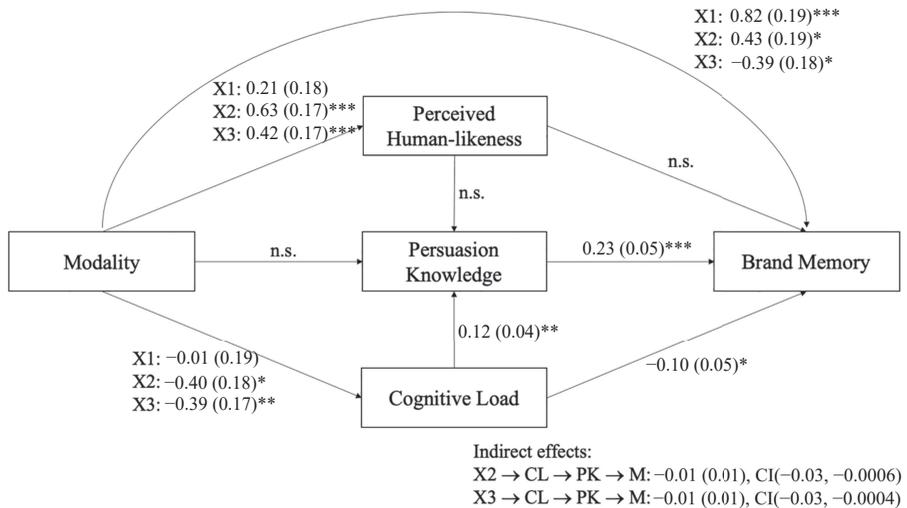
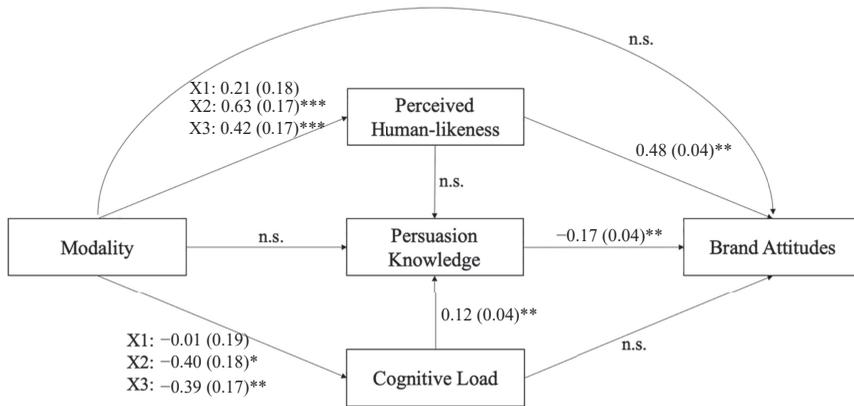


Figure 3.
Model explaining
brand memory

Note(s): X1 = voice vs. voice + interaction history; X2 = voice vs. text; X3 = voice + interaction history vs. text; CL = Cognitive Load; PK = Persuasion Knowledge; M = Brand Memory; * p -value < 0.05, ** p -value < 0.01, *** p -value < 0.001, n.s. = non-significant; standard errors in parentheses

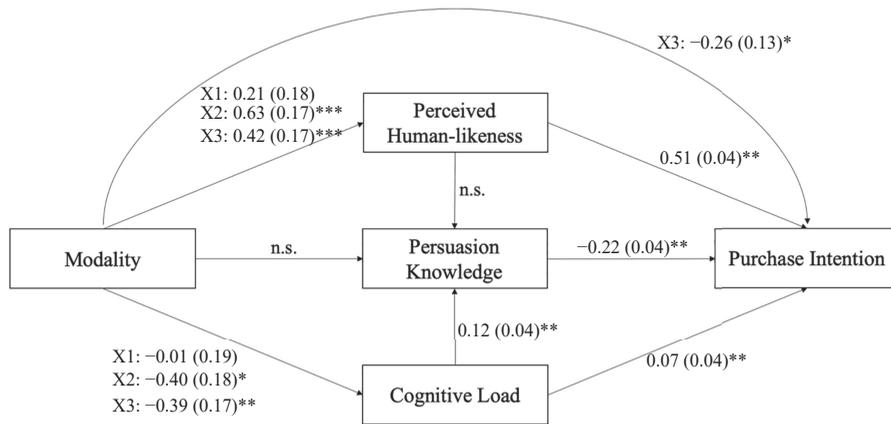


Indirect effects:

X2 → HL → A: 0.30 (0.09), CI (0.14, 0.48) X2 → CL → PK → A: 0.01 (0.01), CI (0.0002, 0.02)
 X3 → HL → A: 0.20 (0.09), CI (0.03, 0.39) X3 → CL → PK → A: 0.01 (0.01), CI (0.0002, 0.02)

Note(s): X1 = voice vs. voice + interaction history; X2 = voice vs. text; X3 = voice + interaction history vs. text; HL = Human-likeness; CL = Cognitive Load; PK = Persuasion Knowledge; A = Brand Attitudes; **p*-value < 0.05, ***p*-value < 0.01, ****p*-value < 0.001, n.s. = non-significant; standard errors in parentheses

Figure 4. Model explaining brand attitudes



Indirect effects:

X2 → HL → PI: 0.32 (0.09), CI (0.15, 0.50) X2 → CL → PK → PI: 0.01 (0.01), CI (0.0003, 0.03)
 X3 → HL → PI: 0.22 (0.09), CI (0.04, 0.41) X3 → CL → PK → PI: 0.01 (0.01), CI (0.0003, 0.03)

Note(s): X1 = voice vs. voice + interaction history; X2 = voice vs. text; X3 = voice + interaction history vs. text; HL = Human-likeness; CL = Cognitive Load; PK = Persuasion Knowledge; PI = Purchase Intention; **p*-value < 0.05, ***p*-value < 0.01, ****p*-value < 0.001; n.s. = non-significant; standard errors in parentheses

Figure 5. Model explaining purchase intention

presented above, perceived human-likeness was significantly higher in the text condition than in the voice only and in the voice + interaction history condition. Furthermore, the path from modality to cognitive load was significant for comparing voice only with text only, and surprisingly, also for comparing voice + interaction history condition and text only condition. In other words, participants in the text only condition invested less mental effort in the interaction than in the other two conditions.

Contrary to expectations, results indicate further that cognitive load positively influences persuasion knowledge ($b = 0.12$, $SE = 0.04$, $p = 0.003$), meaning that the more mental effort participants invested in the conversation the more likely they were to identify a persuasive or selling attempt. Furthermore, we find a small significant negative indirect effect for modality (voice only vs text only, and voice only vs voice + interaction history) on persuasion knowledge mediated by cognitive load. These indirect effects translate into all three advertising effectiveness variables, brand memory (voice only vs text only: $b = -0.01$, $SE = 0.01$, 95% CI $[-0.03, -0.0006]$; voice + interaction history vs text only: $b = -0.01$, $SE = 0.01$, 95% CI $[-0.03, -0.0004]$); brand attitudes (voice only vs text only: $b = -0.01$, $SE = 0.01$, 95% CI $[0.0002, 0.02]$; voice + interaction history vs text only: $b = -0.01$, $SE = 0.01$, 95% CI $[0.0002, 0.02]$); purchase intention (voice only vs text only: $b = -0.01$, $SE = 0.01$, 95% CI $[0.0003, 0.03]$; voice + interaction history vs text only: $b = -0.01$, $SE = 0.01$, 95% CI $[0.0003, 0.03]$).

We do not find any indirect effects of perceived human-likeness on persuasion knowledge, and the three advertising effectiveness variables. In response to the research questions 1–3, interacting with a voice-based virtual assistant does not directly lead to lower persuasion knowledge in comparison to interacting with a virtual assistant that uses a visual display of text (either text only, or voice accompanied by text), nor are these effects mediated by higher perceptions of the virtual assistant's human-likeness. However, cognitive load mediated the effect of modality on persuasion knowledge in an unexpected direction. The higher cognitive load imposed in the voice condition compared to the two conditions displaying text positively influences persuasion knowledge and subsequently advertising effectiveness.

Discussion and conclusion

Virtual assistants have become increasingly important for businesses as a new way to disseminate persuasive messages that are more subtle than traditional forms of advertising. Importantly, virtual assistants can not only be text-based, but are more often based on voice-interactions that can potentially give the assistant a more human touch and make it more persuasive. Hence, the current study examines the persuasive potential of different virtual assistant modalities.

This study adds to the emergent research stream considering virtual assistants in explicitly exploring modality differences between voice and text (and a combination of the two) and provides insights into the effects of persuasion coming from virtual assistants. Drawing on previous research in the field of human–machine communication, the current study extends this work in investigating whether voice in itself (vs text, or a combination of both) influences the perceptions of human-likeness. It contributes to marketing communication theories by applying the persuasion knowledge model (Friestad and Wright, 1994). By doing so, this study includes cognitive load as well as persuasion knowledge and examines consumers' understanding of persuasive intents coming from virtual assistants and downstream effects on brand memory, brand attitudes and purchase intention. Given the increased application of virtual assistants for commercial purposes (e.g. product recommendations), our work focuses on the persuasion process and connects human–machine communication to marketing communication theories. Four main conclusions emerge from this study and have theoretical implications.

Firstly, we find the text-based virtual assistant to be perceived as more human-like (or less machine-like, considering that mean scores are only slightly above midpoint) compared to the voice-based assistant. This effect exists irrespectively of whether the interaction was visually displayed by means of showing the interaction history or not. This finding is contradictory to our expectations formulated in H1 and suggestions of previous research (Cho *et al.*, 2019; Schroeder and Epley, 2016), as voice as used in this experiment decreased the virtual assistant's human-like appeal. This finding greatly contributes to the emerging research stream on human-likeness perceptions of virtual assistants by providing some contradictory results that can spark future research on possible boundary conditions that must be explored.

One possible explanation for this finding can be provided by social information processing (SIP) theory. Based on SIP theory, users rely on available cues to form an impression and a relationship with their interaction partner (Walther, 1992; Walther *et al.*, 2015). For the interaction with a virtual assistant this implies that consumers try to interpret all human-like cues given by the assistant, including speech. In the text only condition, only limited cues are available. This leaves room for consumers' own interpretation, as they have no non-verbal cues available. In the voice condition however, the (synthetic) voice as used in the experiment might have functioned as a cue that created perceptions of machine-likeness. It might have made the non-human nature of the communication partner more obvious.

To relate this finding to the opposing results of previous work by Cho *et al.* (2019) it is worth noting that they found a mediating effect of perceived human-likeness on attitudes toward the virtual assistant for utilitarian, but not for hedonic tasks. We do not know whether the participants in our more specific persuasive scenario experienced the interaction about a dinner recipe recommendation as rather hedonic, or utilitarian. Hence, we are not able to directly compare these results. However, taking the two studies into account can give an indication for possible task-specific moderating factors such as task involvement or preference. Future research should include these and other variables that might explain the different results.

Secondly, the current study contributes to marketing communication research by applying the persuasion knowledge model to examine the persuasive effects of voice (vs text) in virtual assistants. As virtual assistants are increasingly used for persuasive purposes, it is imperative to understand whether it is possible to make conversations more human-like by implementing a certain modality, and whether that translates into persuasiveness. Our findings regarding RQ1 are mixed. We find that modality does not directly influence persuasion knowledge, nor is this effect mediated by perceptions of human-likeness. In other words, not only is a text-based interaction perceived as more human-like than a voice-based interaction in this study, but a perception of human-likeness also does not influence whether consumers identify a persuasive or selling attempt. However, while we do not find any direct effects of modality on persuasion knowledge, we see in our additional analyses that the text-based assistant is perceived as more human-like and positively influences brand attitudes and purchase intention. More research is needed to fully understand the persuasive potential of virtual assistants.

Thirdly, we make a theoretical contribution by including cognitive load as an alternative explanation for persuasiveness. We expected that interacting via voice is more demanding for consumers and increases cognitive load (H2; Berry *et al.*, 2005; Van Zant and Berger, 2019), which in turn also leads them to be less likely to activate their persuasion knowledge (RQ2 and RQ3; Campbell and Kirmani, 2000). Contrary to expectations, our results show that cognitive load did not suppress, but increased persuasion knowledge. This suggests that a task that is more demanding for a user, might make them more alert toward the content of the interaction. Moreover, even though post hoc differences are not significant for our ANCOVA to test H2 and must be handled with caution, we do see effects of modality on cognitive load

when testing the full model. Cognitive load was lower when communicating with the text-based virtual assistant compared to communicating with one of the two voice-based assistants.

This is surprising, considering our third condition of a voice assistant accompanied by the display of interaction history. We find that showing an interaction history did not lift cognitive load. Quite on the contrary, voice accompanied by interaction history induced more cognitive load than text alone. If the text-output would have been the driver for cognitive load, we would have seen differences between communicating via voice only compared to a modality that has a visual display of text (so either text only, or voice with interaction history), but no differences between the two conditions that included text. This leads us to the conclusion that the differences in cognitive load of voice- and text-based virtual assistant interaction could be attributed to the input modality instead. In the two voice-conditions, participants used voice as input modality to interact with the virtual assistant, which might have led to differences in cognitive load.

It must be noted, however, that the relative amount of cognitive load is below the scale's midpoint, hence relatively low. One explanation could be that the interaction about a dinner recommendation as chosen in this experiment is a relatively easy topic to engage in. Additional research is needed to examine whether cognitive load is influenced by the virtual assistant's (input as well as output) modality in the context of more demanding topics.

Lastly, we contribute to knowledge on the overall effectiveness of different virtual assistant modalities. We studied whether persuasion knowledge translates into persuasive outcomes for virtual assistants. Based on previous research on the effects of persuasion knowledge (e.g. Boerman and van Reijmersdal, 2020; Boerman *et al.*, 2017; Obermiller *et al.*, 2005; Tutaj and van Reijmersdal, 2012; van Reijmersdal *et al.*, 2016), we proposed in H3 that persuasion knowledge is positively related to brand memory, and negatively related to brand attitudes and purchase intention. Notwithstanding the lack of effects of virtual assistant modality on persuasion knowledge, our study confirms these expectations. It thereby shows that findings from these research lines can be translated to the context of virtual assistants. Hence, we can show and corroborate previous research (Voorveld and Araujo, 2020) that using the persuasion knowledge model is generally a useful tool to explain persuasive attempts coming from a virtual assistant. Notably, the positive effects of persuasion knowledge on brand memory were stronger than the negative effects on brand attitudes and purchase intention.

Managerial implications

Our findings have managerial implications. When concerning emerging technologies it is often assumed that a higher degree of "human touch" can lead to more positive affective evaluations (see, e.g. Sivaramakrishnan *et al.*, 2007). Our study can confirm and extend this notion by showing that perceived human-likeness can positively influence persuasive outcomes such as brand attitudes and purchase intention. For businesses, it might therefore be valuable to invest in emerging technologies that provide a human appeal in consumer-brand interactions. However, simply using voice as a presumably human cue might not suffice. Quite on the contrary, voice can be experienced as less human-like than text by consumers, which implies that businesses must carefully evaluate how human-likeness can be conveyed.

Furthermore, reflecting on the differences of voice- and text-based virtual assistant interactions, our findings can give a first indication that voice as a communication modality can *increase* persuasion knowledge by being cognitively more demanding. For the development and implementation of virtual assistants in business, it might therefore be beneficial to actively engage consumers. Persuasion knowledge in the context of virtual assistants makes consumers aware of the commercial nature of an interaction, but at the same

time also positively influences cognitive outcomes such as brand memory. As we could show, consumers' brand memory in general is relatively low, hence actively engaging consumers might help businesses to strengthen the visibility of their brands. Our findings further indicate that a combination of voice and text might be most successful to do so.

Societal implications

Furthermore, our findings have societal implications and inform regulators and policymakers in finding ways to empower users of virtual assistants. Based on the findings of our study, voice used as a human-like cue does not disallow consumers to cope with persuasive attempts. However, we also confirm direct effects of perceived human-likeness on more affective outcomes such as brand attitudes and purchase intention. Considering the *threats* of increasingly human-like interfaces to influence and possibly mislead consumers, we suggest to not only explicitly inform recipients about the commercial nature of the conversation (as done with sponsorship disclosure, e.g. [Boerman et al., 2017](#)) but also to inform them about the non-human nature of the interaction.

Moreover, we find that persuasion knowledge as a cognitive response is influenced by the amount of cognitive load used when interacting with an assistant. The more engaged consumers are in a conversation, the more likely they might be to recognize a persuasive attempt. Hence, policymakers should try to not only inform, but involve consumers in interactions with technologies such as conversational agents.

Limitations and suggestions for future research

This study merits mentioning two methodological limitations that bring suggestions for future research. Firstly, even though randomly assigned to conditions, more participants interacted with the text-based virtual assistant compared to the two voice-based assistants, suggesting that more participants dropped out in the voice only and voice + interaction history condition compared to the text condition. Participants may have experienced some sort of hindrance interacting with the voice-based virtual assistant, especially when the interaction was not visually displayed in text. However, to be able to measure persuasion knowledge and brand memory, it was necessary to only include participants that had a full interaction with the virtual assistant. Hence, future research should consider possible difficulties for participants to interact with voice only. Furthermore, we see that participants that communicated with the text-based assistant indicated that they were more familiar with virtual assistants in general. Even though we were able to control for familiarity in our analysis, we suggest future research to include familiarity with voice-based communication and familiarity with text-based communication as two separate constructs.

Secondly, we decided to use a younger male voice for this experiment. We believe that this created a realistic scenario since we chose a voice readily available on Google Chrome. However, future research should further examine the persuasive potential and expand the growing body of research on different auditory cues such as gender, emotionality or vocal pitch. Furthermore, we decided to mirror input and output modality in this study, meaning that participants used voice to interact with the two voice-based assistants (with and without interaction history displayed), and text to interact with the text-based assistant. This is as also suggested by previous scholars ([Cho et al., 2019](#)) and adds naturalness to the interaction, since it resembles how virtual assistants are used in practice (e.g. Google Assistant, Amazon Alexa). However, since we suggest that cognitive load of voice- and text-based virtual assistant interaction might be attributed to the input modality, it is advised for future research to add another layer of complexity and experimentally manipulate input modality in addition to output modality. This will enable an even more fine-grained analysis of the interplay between the two.

Overall conclusion and contribution

In sum, our study shows that a text-based virtual assistant is perceived as more human-like compared to a voice-based assistant (regardless of whether the interaction history is displayed). By providing contradictory findings to previous literature, we add to the field on human-machine communication and point toward the need to further extend the body of research explicitly comparing modality differences in their perceptions of human-likeness. We furthermore contribute to the understanding of the concept of human-likeness of technologies with persuasive purposes by showing that perceived human-likeness positively influences brand attitudes and purchase intention. Moreover, we extend modality research to the context of virtual assistants by demonstrating that communicating via voice is higher in cognitive load than communicating via text. Lastly, we contribute to research on modality and persuasion knowledge by showing that a persuasive or selling intent is easier recognized in a voice-based interaction (by being cognitively more demanding) in comparison to text.

Notes

1. Note that H2 and RQ2 divert from the pre-registration. In accordance with Cognitive Load Theory, we decided to consistently adopt the concept name “cognitive load” instead of “cognitive processing” throughout the manuscript.
2. <https://osf.io/yrzgb/>
3. The continuous variables (perceived human-likeness, cognitive load, persuasion knowledge, attitudes, purchase intention and number of correctly identified branded items in recognition task) were screened by using Mahalanobis distance. Based on the chi-square distribution ($df = 6$, $p = 0.001$), the cutoff point is 22.46.
4. Note that the sample sizes are unequal, which implies that violation of assumptions of normality and homogeneity of variances can be severe (Field *et al.*, 2012). However, since assumptions were met, we proceeded with a parametric test and conducted ANCOVA.
5. <https://osf.io/yrzgb/>
6. We decided for a unidimensional scale, as Paas *et al.* (2003, p. 66) state “[they] are sensitive to relatively small differences in cognitive load and that they are [a] valid, reliable and unintrusive” measurement of cognitive load.
7. Note that Levene’s test indicated unequal variances ($F(2, 447) = 6.24$, $p = 0.002$).

References

- Accenture (2018), “Chatbots are here to stay. So what are you waiting for?”, available at: https://www.accenture.com/_acnmedia/pdf-77/accenture-research-conversational-ai-platforms.pdf (accessed 30 January 2020).
- Araujo, T. (2020), “Conversational agent research toolkit: an alternative for creating and managing chatbots for experimental research”, *Computational Communication Research*, Vol. 2 No. 1, pp. 35-51.
- Bartneck, C., Kulić, D., Croft, E. and Zoghbi, S. (2009), “Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots”, *International Journal of Social Robotics*, Vol. 1 No. 1, pp. 71-81.
- Berger, J. and Iyengar, R. (2013), “Communication channels and word of mouth: how the medium shapes the message”, *Journal of Consumer Research*, Vol. 40 No. 3, pp. 567-579.
- Berry, D.C., Butler, L.T. and De Rosis, F. (2005), “Evaluating a realistic agent in an advice-giving task”, *International Journal of Human Computer Studies*, Vol. 63 No. 3, pp. 304-327.

- Besik, H. (2019), "91% of brands are investing in voice: how to make it work", available at: <https://theblog.adobe.com/91-of-brands-are-investing-in-voice-how-to-make-it-work/> (accessed 30 January 2020).
- Boerman, S.C. and van Reijmersdal, E.A. (2020), "Disclosing influencer marketing on YouTube to children: the moderating role of para-social relationship", *Frontiers in Psychology*, Vol. 10 January, pp. 1-15.
- Boerman, S.C., van Reijmersdal, E.A. and Neijens, P.C. (2012), "Sponsorship disclosure: effects of duration on persuasion knowledge and brand responses", *Journal of Communication*, Vol. 62 No. 6, pp. 1047-1064.
- Boerman, S.C., Willemsen, L.M. and Van Der Aa, E.P. (2017), "'This post is sponsored': effects of sponsorship disclosure on persuasion knowledge and electronic word of mouth in the context of Facebook", *Journal of Interactive Marketing*, Vol. 38, pp. 82-92.
- Boerman, S.C., van Reijmersdal, E.A., Rozendaal, E. and Dima, A.L. (2018), "Development of the persuasion knowledge scales of sponsored content (PKS-SC)", *International Journal of Advertising*, Vol. 37 No. 5, pp. 671-697.
- Brehm, J.W. (1966), *A Theory of Psychological Reactance*, Academic Press, New York.
- Buijzen, M., Van Reijmersdal, E.A. and Owen, L.H. (2010), "Introducing the PCMC model: an investigative framework for young people's processing of commercialized media content", *Communication Theory*, Vol. 20 No. 4, pp. 427-450.
- Campbell, M.C. (1995), "When attention-getting advertising tactics elicit consumer inferences of manipulative intent: the importance of balancing benefits and investments", *Journal of Consumer Psychology*, Vol. 4 No. 3, pp. 225-254.
- Campbell, M.C. and Kirmani, A. (2000), "Consumers' use of persuasion knowledge: the effects of accessibility and cognitive capacity on perceptions of an influence agent", *Journal of Consumer Research*, Vol. 27 No. 1, pp. 69-83.
- Cheng, X., Bao, Y., Zarifis, A., Gong, W. and Mou, J. (2021), "Exploring consumers' response to text-based chatbots in e-commerce: the moderating role of task complexity and chatbot disclosure", *Internet Research*, Vol. 32 No. 2, pp. 496-517.
- Cho, E., Molina, M.D. and Wang, J. (2019), "The effects of modality, device, and task differences on perceived human likeness of voice-activated virtual assistants", *Cyberpsychology, Behavior and Social Networking*, Vol. 22 No. 8, pp. 515-520.
- Choi, D., Bang, H., Wojdyski, B.W., Lee, Y.I. and Keib, K.M. (2018), "How brand disclosure timing and brand prominence influence consumer's intention to share branded entertainment content", *Journal of Interactive Marketing*, Vol. 42, pp. 18-31.
- Cox, L. (2018), "5 business uses of voice based virtual assistants", available at: <https://disruptionhub.com/5-business-uses-of-voice-based-virtual-assistants/> (accessed 30 January 2020).
- Dabholkar, P.A. and Sheng, X. (2012), "Consumer participation in using online recommendation agents: effects on satisfaction, trust, and purchase intentions", *Service Industries Journal*, Vol. 32 No. 9, pp. 1433-1449.
- Davis, F.D., Bagozzi, R.P. and Warshaw, P.R. (1989), "User acceptance of computer technology: a comparison of two theoretical models authors", *Management Science*, Vol. 35 No. 8, pp. 982-1003.
- de Visser, E.J., Monfort, S.S., McKendrick, R., Smith, M.A.B., McKnight, P.E., Krueger, F. and Parasuraman, R. (2016), "Almost human: anthropomorphism increases trust resilience in cognitive agents", *Journal of Experimental Psychology: Applied*, Vol. 22 No. 3, pp. 331-349.
- Ding, Y., Prepin, K., Huang, J., Pelachaud, C. and Artieres, T. (2014), "Laughter animation synthesis", *Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2014)*, pp. 773-780.
- Duggal, R. (2020), "Is voice the next big thing to transform consumer behavior?", available at: <https://www.forbes.com/sites/forbescommunicationscouncil/2020/03/23/is-voice-the-next-big-thing-to-transform-consumer-behavior/> (accessed 28 October 2020).

- Epley, N., Waytz, A. and Cacioppo, J.T. (2007), "On seeing human: a three-factor theory of anthropomorphism", *Psychological Review*, Vol. 114 No. 4, pp. 864-886.
- Feine, J., Gnewuch, U., Morana, S. and Maedche, A. (2019), "A taxonomy of social cues for conversational agents", *International Journal of Human-Computer Studies*, Vol. 132 July, pp. 138-161.
- Field, A.P., Miles, J. and Field, Z. (2012), *Discovering Statistics Using R*, SAGE Publications, London.
- Forlizzi, J., Zimmerman, J., Mancuso, V. and Kwak, S. (2007), "How interface agents affect interaction between humans and computers", *Proceedings of the 2007 Conference on Designing Pleasurable Products and Interfaces, DPPI'07*, pp. 209-221.
- Fransen, M.L., Verlegh, P.W.J., Kirmani, A. and Smit, E.G. (2015), "A typology of consumer strategies for resisting advertising, and a review of mechanisms for countering them", *International Journal of Advertising*, Vol. 34 No. 1, pp. 6-16.
- Friestad, M. and Wright, P. (1994), "The persuasion knowledge model: how people cope with persuasion attempts", *Journal of Consumer Research*, Vol. 21 No. 1, pp. 1-31.
- Gefen, D. and Straub, D.W. (2004), "Consumer trust in B2C e-Commerce and the importance of social presence: experiments in e-Products and e-Services", *Omega*, Vol. 32 No. 6, pp. 407-424.
- Grand View Research Inc (2021), "Chatbot market size worth \$2,485.7 million by 2028 | CAGR: 24.9%", available at: <https://www.grandviewresearch.com/press-release/global-chatbot-market> (accessed 15 November 202).
- Gray, P. (2016), "The rise of intelligent virtual assistants", available at: <https://www.interactions.com/blog/intelligent-virtual-assistant/rise-intelligent-virtual-assistants/> (accessed 20 May 2020).
- Guzman, A.L. and Lewis, S.C. (2020), "Artificial intelligence and communication: a Human-Machine Communication research agenda", *New Media and Society*, Vol. 22 No. 1, pp. 70-86.
- Ham, C.D., Nelson, M.R. and Das, S. (2015), "How to measure persuasion knowledge", *International Journal of Advertising*, Vol. 34 No. 1, pp. 17-53.
- Hayes, A.F. (2017), *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-based Approach*, The Guilford press, New York.
- Ho, C.C. and MacDorman, K.F. (2010), "Revisiting the uncanny valley theory: developing and validating an alternative to the Godspeed indices", *Computers in Human Behavior*, Vol. 26 No. 6, pp. 1508-1518.
- Hossain, M.T. and Saini, R. (2014), "Suckers in the morning, skeptics in the evening: time-of-day effects on consumers' vigilance against manipulation", *Marketing Letters*, Vol. 25 No. 2, pp. 109-121.
- Kim, Y. and Sundar, S.S. (2012), "Anthropomorphism of computers: is it mindful or mindless?", *Computers in Human Behavior*, Vol. 28 No. 1, pp. 241-250.
- Kirmani, A. and Campbell, M.C. (2008), "I know what you're doing and why you're doing it: the use of persuasion knowledge model in consumer research", in Haugtvedt, C.P., Herr, P.M. and Kardes, F.R. (Eds), *Handbook of Consumer Psychology*, Taylor & Francis Group, New York, pp. 549-574.
- Le Bigot, L., Jamet, E. and Rouet, J.F. (2004), "Searching information with a natural language dialogue system: a comparison of spoken vs written modalities", *Applied Ergonomics*, Vol. 35 No. 6, pp. 557-564.
- Leahy, W. and Sweller, J. (2011), "Cognitive load theory and the effects of transient information on the modality effect", *Applied Cognitive Psychology*, Vol. 25, pp. 943-951.
- Lee, K.M. (2004), "Presence, explicated", *Communication Theory*, Vol. 14 No. 1, pp. 27-50.
- Lee, E.-J. (2010), "What triggers social responses to flattering computers? Experimental tests of anthropomorphism and mindlessness explanations", *Communication Research*, Vol. 37 No. 2, pp. 191-214.
- Lee, K.M. and Nass, C. (2003), "Designing social presence of social actors in human computer interaction", *Proceedings of the Conference on Human Factors in Computing Systems - CHI '03*, No. 5, p. 289.

- Lee, C.T., Pan, L.Y. and Hsieh, S.H. (2021), "Artificial intelligent chatbots as brand promoters: a two-stage structural equation modeling-artificial neural network approach", *Internet Research*, Vol. 32 No. 4, pp. 1329-1356.
- Louwerse, M.M., Graesser, A.C., Lu, S. and Mitchell, H.H. (2005), "Social cues in animated conversational agents", *Applied Cognitive Psychology*, Vol. 19 No. 6, pp. 693-704.
- Moriuchi, E. (2019), "Okay, Google!: an empirical study on voice assistants on consumer engagement and loyalty", *Psychology and Marketing*, Vol. 36 No. 5, pp. 489-501.
- Nowak, K.L. and Biocca, F. (2003), "The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments", *Presence: Teleoperators and Virtual Environments*, Vol. 12 No. 5, pp. 481-494.
- Obermiller, C., Spangenberg, E. and MacLachlan, D.L. (2005), "Ad skepticism: the consequences of disbelief", *Journal of Advertising*, Vol. 34 No. 3, pp. 7-17.
- Paas, F. (1992), "Training strategies for attaining transfer of problem-solving skill in statistics: a cognitive-learning approach", *Journal of Educational Psychology*, Vol. 84 No. 4, pp. 429-434.
- Paas, F. and Van Merriënboer, J.J.G. (1994), "Instructional control of cognitive load in the training of complex cognitive tasks", *Educational Psychology Review*, Vol. 6 No. 4, pp. 351-371.
- Paas, F., Tuovinen, J.E., Tabbers, H. and van Gerven, P.W.M. (2003), "Cognitive load measurement as a means to advance cognitive load theory", *Educational Psychologist*, Vol. 38 No. 1, pp. 63-71.
- Paivio, A. (1986), *Mental Representations: A Dual-Coding Approach*, Oxford University Press, New York.
- Pastore, R. (2014), "Multimedia: learner preferences for multimedia learning", *Journal of Multimedia Processing and Technologies*, Vol. 5 No. 4, pp. 134-142.
- Pfau, M. (1990), "A channel approach to television influence", *Journal of Broadcasting & Electronic Media*, Vol. 34 No. 2, pp. 195-214.
- Pinker, S. and Bloom, P. (1990), "Natural language and natural selection", *Behavioral and Brain Sciences*, Vol. 13, pp. 707-727.
- Rabideau, C. (2018), "8 Things you didn't know Amazon Alexa could do in the kitchen", available at: <https://www.reviewed.com/smarthome/features/8-things-you-didnt-know-amazon-alexa-could-do-in-the-kitchen> (accessed 22 October 2020).
- Rapp, A., Curti, L., Boldi, A., Curti, L. and Boldi, A. (2021), "The human side of human-chatbot interaction: a systematic literature review of ten years of research on text-based chatbots", *Human-Computer Studies*, Vol. 151, pp. 1-24.
- Reeves, B. and Nass, C. (1996), *The Media Equation: How People Treat Computers, Television, and New Media like Real People and Places*, University Press, Cambridge.
- Research and Markets (2021), "Global chatbot market 2021-2025", available at: <https://www.researchandmarkets.com/reports/5398481/global-chatbot-market-2021-2025> (accessed 15 November 2021).
- Rhee, C.E. and Choi, J. (2020), "Effects of personalization and social role in voice shopping: an experimental study on product recommendation by a conversational voice agent", *Computers in Human Behavior*, Vol. 109, pp. 1-11.
- Schmidt-Weigand, F., Kohnert, A. and Glowalla, U. (2010), "A closer look at split visual attention in system- and self-paced instruction in multimedia learning", *Learning and Instruction*, Vol. 20 No. 2, pp. 100-110.
- Schroeder, J. and Epley, N. (2016), "Mistaking minds and machines: how speech affects dehumanization and anthropomorphism", *Journal of Experimental Psychology: General*, Vol. 145 No. 11, pp. 1427-1437.
- Seifert, D. (2020), "The best smart display to buy right now", available at: <https://www.theverge.com/2020/2/18/21139800/best-smart-display-home-amazon-google-alexa-assistant-echo-nest> (accessed 29 October 2020).

-
- Sivaramakrishnan, S., Fang, W. and Zaiyong, T. (2007), "Giving an 'e-human touch' to e-tailing: the moderating roles of static information quantity and consumption motive in the effectiveness of an anthropomorphic information agent", *Journal of Interactive Marketing*, Vol. 21 No. 1, pp. 60-75.
- Tillman, M. and O'Boyle, B. (2019), "What is Google Assistant and what can it do?", available at: <https://www.pocket-lint.com/apps/news/google/137722-what-is-google-assistant-how-does-it-work-and-which-devices-offer-it> (accessed 30 January 2020).
- Tutaj, K. and van Reijmersdal, E.A. (2012), "Effects of online advertising format and persuasion knowledge on audience reactions", *Journal of Marketing Communications*, Vol. 18 No. 1, pp. 5-18.
- van Noort, G., Antheunis, M.L. and van Reijmersdal, E.A. (2012), "Social connections and the persuasiveness of viral campaigns in social network sites: persuasive intent as the underlying mechanism", *Journal of Marketing Communications*, Vol. 18 No. 1, pp. 39-53.
- Van Reijmersdal, E.A., Rozendaal, E. and Buijzen, M. (2012), "Effects of prominence, involvement, and persuasion knowledge on children's cognitive and affective responses to advergames", *Journal of Interactive Marketing*, Vol. 26 No. 1, pp. 33-42.
- van Reijmersdal, E.A., Franssen, M.L., van Noort, G., Oprea, S.J., Vandenberg, L., Reusch, S., van Lieshout, F. and Boerman, S.C. (2016), "Effects of disclosing sponsored content in blogs", *American Behavioral Scientist*, Vol. 60 No. 12, pp. 1458-1474.
- Van Zant, A.B. and Berger, J. (2019), "How the voice persuades", *Journal of Personality and Social Psychology*, Vol. 118 No. 4, pp. 1-22.
- Voorveld, H.A.M. and Araujo, T. (2020), "How social cues in virtual assistants influence concerns and persuasion: the role of voice and a human name", *Cyberpsychology, Behavior, and Social Networking*, Vol. 23 No. 10, pp. 689-696.
- Walther, J.B. (1992), "Interpersonal effects in computer-mediated interaction: a relational perspective", *Communication Research*, Vol. 19 No. 1, pp. 52-90.
- Walther, J.B., Van Der Heide, B., Ramirez, A., Burgoon, J.K. and Peña, J. (2015), "Interpersonal and hyperpersonal dimensions of computer-mediated communication", in Sundar, S.S. (Ed.), *The Handbook of the Psychology of Communication Technology*, Wiley Blackwell, New York, pp. 3-24.
- Zhou, L., Yang, Z. and Hui, M.K. (2010), "Non-local or local brands? A multi-level investigation into confidence in brand origin identification and its strategic implications", *Journal of the Academy of Marketing Science*, Vol. 38 No. 2, pp. 202-218.

Appendix

Is voice really persuasive?

	Text (n = 203)	M (SD) Voice (n = 113)	Voice + IH (n = 134)
Perceived human-likeness	4.20 (1.50)	3.59 (1.32)	3.84 (1.44)
Cognitive load	3.11 (1.47)	3.49 (1.40)	3.41 (1.52)
Persuasion knowledge	4.55 (1.38)	4.29 (1.13)	4.54 (1.26)
Brand memory	2.36 (1.57)	1.61 (1.33)	2.66 (1.66)
Brand attitudes	5.00 (1.44)	4.84 (1.10)	5.05 (1.21)
Purchase intention	4.19 (1.41)	4.21 (1.28)	4.30 (1.29)
Age	45.77 (16.13)	45.81 (15.60)	47.02 (15.70)
Familiarity assistant	4.56 (1.40)	3.57 (1.53)	4.07 (1.56)
Familiarity brand	4.37 (1.52)	4.02 (1.48)	4.70 (1.37)
Preference voice/text	5.69 (1.81)	4.64 (2.10)	4.08 (2.06)

Note(s): IH = Interaction History

423

Table A1.
Descriptive statistics

	Coefficient (SE)	t	p
<i>Outcome variable: Perceived human-likeness</i>			
X1: voice vs voice + IH	0.21 (0.18)	1.12	0.263
X2: voice vs text	0.63 (0.17)	3.59	0.000
X3: voice + IH vs text	0.42 (0.17)	2.52	0.012
<i>Outcome variable: Cognitive load</i>			
X1: voice vs voice + IH	-0.01 (0.19)	-0.03	0.979
X2: voice vs text	-0.40 (0.18)	-2.18	0.030
X3: voice + IH vs text	-0.39 (0.17)	-2.24	0.010
<i>Outcome variable: Persuasion knowledge</i>			
X1: voice vs voice + IH	0.15 (0.17)	0.91	0.364
X2: voice vs text	0.14 (0.16)	0.87	0.384
X3: voice + IH vs text	-0.01 (0.15)	-0.07	0.945
Perceived Human-likeness	0.04 (0.04)	0.86	0.393
Cognitive Load	0.12 (0.04)	2.96	0.003
<i>Outcome variable: Brand memory</i>			
X1: voice vs voice + IH	0.82 (0.19)	4.30	0.000
X2: voice vs text	0.43 (0.19)	2.31	0.021
X3: voice + IH vs text	-0.39 (0.18)	-2.23	0.027
Perceived Human-likeness	0.03 (0.05)	0.51	0.609
Cognitive Load	-0.10 (0.05)	-2.09	0.038
Persuasion Knowledge	0.23 (0.05)	4.27	0.000

Indirect effects

	Effect (SE)	Confidence interval Lower limit	Upper limit
X1 → HL → M	0.01 (0.01)	-0.02	0.04
X2 → HL → M	0.02 (0.03)	-0.05	0.09
X3 → HL → M	0.01 (0.02)	-0.04	0.06
X1 → CL → M	0.00 (0.02)	-0.04	0.05

(continued)

Table A2.
Full model explaining brand memory

<i>Indirect effects</i>			
	Effect (SE)	Confidence interval Lower limit	Upper limit
X2 → CL → M	0.04 (0.03)	-0.002	0.11
X3 → CL → M	0.04 (0.03)	-0.003	0.10
X1 → PK → M	0.04 (0.04)	-0.04	0.12
X2 → PK → M	0.03 (0.04)	-0.04	0.11
X3 → PK → M	-0.003 (0.04)	-0.08	0.07
X1 → HL → PK → M	0.00 (0.00)	-0.004	0.01
X2 → HL → PK → M	0.01 (0.01)	-0.01	0.02
X3 → HL → PK → M	0.004 (0.01)	-0.01	0.02
X1 → CL → PK → M	0.00 (0.01)	-0.01	0.01
X2 → CL → PK → M	-0.01 (0.01)	-0.03	-0.0006
X3 → CL → PK → M	-0.01 (0.01)	-0.03	-0.0004

Note(s): Perceived Human-likeness, $F(5, 444) = 8.90, p < 0.001, R^2 = 0.09$; Cognitive Load, $F(5, 444) = 2.73, p = 0.019, R^2 = 0.03$; Persuasion Knowledge, $F(7, 442) = 3.44, p = 0.001, R^2 = 0.05$; Brand Memory, $F(8, 441) = 11.78, p < 0.001, R^2 = 0.18$; controlled for brand familiarity, familiarity with virtual assistants, modality preference; IH = Interaction History; HL = Perceived Human-likeness; CL = Cognitive Load; PK = Persuasion Knowledge; M = Brand Memory

Table A2.

	Coefficient (SE)	<i>t</i>	<i>p</i>
<i>Outcome variable: Brand Attitudes</i>			
X1: voice vs voice + IH	0.08 (0.14)	0.60	0.549
X2: voice vs text	-0.16 (0.14)	-1.18	0.237
X3: voice + IH vs text	-0.24 (0.13)	-1.90	0.059
Perceived Human-likeness	0.48 (0.04)	13.33	0.000
Cognitive Load	0.06 (0.04)	1.66	0.099
Persuasion Knowledge	-0.17 (0.04)	-4.20	0.000

<i>Indirect effects</i>			
	Effect (SE)	Confidence interval Lower limit	Upper limit
X1 → HL → A	0.10 (0.08)	-0.06	0.26
X2 → HL → A	0.30 (0.09)	0.14	0.48
X3 → HL → A	0.20 (0.09)	0.03	0.39
X1 → CL → A	0.00 (0.01)	-0.03	0.03
X2 → CL → A	-0.02 (0.02)	-0.07	0.01
X3 → CL → A	-0.03 (0.02)	-0.07	0.01
X1 → PK → A	-0.03 (0.03)	-0.09	0.03
X2 → PK → A	-0.02 (0.03)	-0.08	0.03
X3 → PK → A	0.002 (0.03)	-0.05	0.05
X1 → HL → PK → A	0.00 (0.00)	-0.01	0.00
X2 → HL → PK → A	0.00 (0.01)	-0.02	0.01
X3 → HL → PK → A	-0.003 (0.00)	-0.01	0.005
X1 → CL → PK → A	0.00 (0.00)	-0.01	0.01
X2 → CL → PK → A	0.01 (0.01)	0.0002	0.02
X3 → CL → PK → A	0.01 (0.01)	0.0002	0.02

Table A3.
Full model explaining brand attitudes

Note(s): $F(8, 441) = 27.91, p < 0.001, R^2 = 0.34$; controlled for brand familiarity, familiarity with virtual assistants, modality preference; IH = Interaction History; HL = Perceived Human-likeness; CL = Cognitive Load; PK = Persuasion Knowledge; A = Brand Attitudes

Is voice really persuasive?

	Coefficient (SE)	<i>t</i>	<i>p</i>
<i>Outcome variable: Purchase intention</i>			
X1: voice vs voice + IH	0.00 (0.14)	0.01	0.994
X2: voice vs text	-0.25 (0.14)	-1.87	0.062
X3: voice + IH vs text	-0.26 (0.13)	-1.97	0.049
Perceived Human-likeness	0.51 (0.04)	14.08	0.000
Cognitive Load	0.07 (0.04)	2.06	0.040
Persuasion Knowledge	-0.22 (0.04)	-5.57	0.000

Indirect effects

	Effect (SE)	Confidence interval	
		Lower limit	Upper limit
X1 → HL → PI	0.10 (0.09)	-0.07	0.29
X2 → HL → PI	0.32 (0.09)	0.15	0.50
X3 → HL → PI	0.22 (0.09)	0.04	0.41
X1 → CL → PI	0.00 (0.02)	-0.03	0.03
X2 → CL → PI	-0.03 (0.02)	-0.09	0.00
X3 → CL → PI	-0.03 (0.02)	-0.07	0.00
X1 → PK → PI	-0.04 (0.04)	-0.11	0.03
X2 → PK → PI	-0.03 (0.03)	-0.10	0.03
X3 → PK → PI	0.002 (0.04)	-0.07	0.07
X1 → HL → PK → PI	0.00 (0.00)	-0.01	0.00
X2 → HL → PK → PI	-0.01 (0.01)	-0.02	0.01
X3 → HL → PK → PI	-0.004 (0.01)	-0.02	0.01
X1 → CL → PK → PI	0.00 (0.01)	-0.01	0.01
X2 → CL → PK → PI	0.01 (0.01)	0.0003	0.03
X3 → CL → PK → PI	0.01 (0.01)	0.0003	0.03

Note(s): $F(8, 441) = 32.89$, $p < 0.001$, $R^2 = 0.37$; controlled for brand familiarity, familiarity with virtual assistants, modality preference; IH = Interaction History; HL = Perceived Human-likeness; CL = Cognitive Load; PK = Persuasion Knowledge; PI = Purchase Intention

Table A4. Full model explaining purchase intention

Corresponding author

Carolin Ischen can be contacted at: c.ischen@uva.nl

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgrouppublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com