Original Article

Is Hearing Really Believing?

The Importance of Modality for Perceived Message
Credibility During Information Search With Smart Speakers

Franziska Gaiser[1] and Sonja Utz[1,2]

[1]Leibniz-Institut für Wissensmedien, Tübingen, Germany
[2]University of Tübingen, Germany

Abstract. Smart speakers are becoming popular all over the world and offer an alternative to conventional web search. We conducted two experiments to investigate whether different message modalities affect credibility perceptions, which role sponsor credibility and message accuracy play, and if this role differs in the two modalities. Based on the MAIN Model by Sundar (2008), we assumed that modality, that is, whether the information is given textually or aurally, can affect credibility assessments. To investigate this, two online experiments with a 2 (modality: smart speaker/search engine) × 2 (message accuracy: high/low) ×3 (sponsor credibility: high/low/none) mixed factorial design were conducted (n = 399 and 398). Information presented by the voice of a smart speaker was generally perceived as more credible. Results further showed that no source attribution and low message accuracy affected message credibility less in the auditive than in the textual modality, especially for participants with less topic involvement. With this, we gained valuable insights into the role of smart speakers for information search and potential downsides of this usage.

Keywords: smart speakers, voice assistants, modality, credibility, search engines

For most, the Internet has become the primary source of information. The popular approach to searching for information online is to use search engines like Google. Recently, smart speakers with voice assistants, also referred to as "voice user interfaces" (Porcheron et al., 2018, p. 1), are becoming popular. By allowing for hands-free control, they offer a convenient alternative to conventional screen-based web search (Zuberer, 2017). According to the latest Smart Audio Report, 24% of Americans reported owning smart speakers (The Smart Audio Report, 2020) and relying on them to answer questions. But this new way of gaining information poses questions "related to the quality, source diversity, and comprehensiveness of information conveyed by these devices" (Dambanemuya & Diakopoulos, 2021, p. 1).

Information search with smart speakers differs from search with search engines. Search engines deliver a text-based list of results to evaluate. Smart speakers usually provide one result read by an artificial voice, making users more dependent on the search algorithm; the other results are invisible (Gruber et al., 2021). The quality of this single

response thus becomes more important (Dambanemuya & Diakopoulos, 2021). *The New York Times* recently compared search results for conspiracy theories on Google and Bing and found that Bing displayed more untrustworthy websites (Thompson, 2022). This result is alarming because Amazon Alexa uses Bing for search queries. It is, therefore, important to examine how critically users examine the single auditive answer from smart speakers.

Moreover, smart speakers often do not mention the source of information. In a study on information quality of Amazon Alexa, Dambanemuya and Diakopoulos (2021) found that most responses were of unknown provenance, which makes it more difficult to determine their trustworthiness (Metzger et al., 2003).

Based on these considerations, the question arises of whether these peculiarities and the fact that the message is delivered aurally alter information processing and evaluation. Do message modality, accuracy, and source provenance affect information search and credibility perceptions? To our knowledge, the credibility of searches with a voice output has not been explicitly compared with text-based web search, although getting the information aurally or the lack of information provenance could influence credibility perceptions (Dambanemuya & Diakopoulos, 2021).

To fill this gap, we investigate how message modality (auditive vs. textual) impacts credibility perception during information search. Additionally, we investigate what role
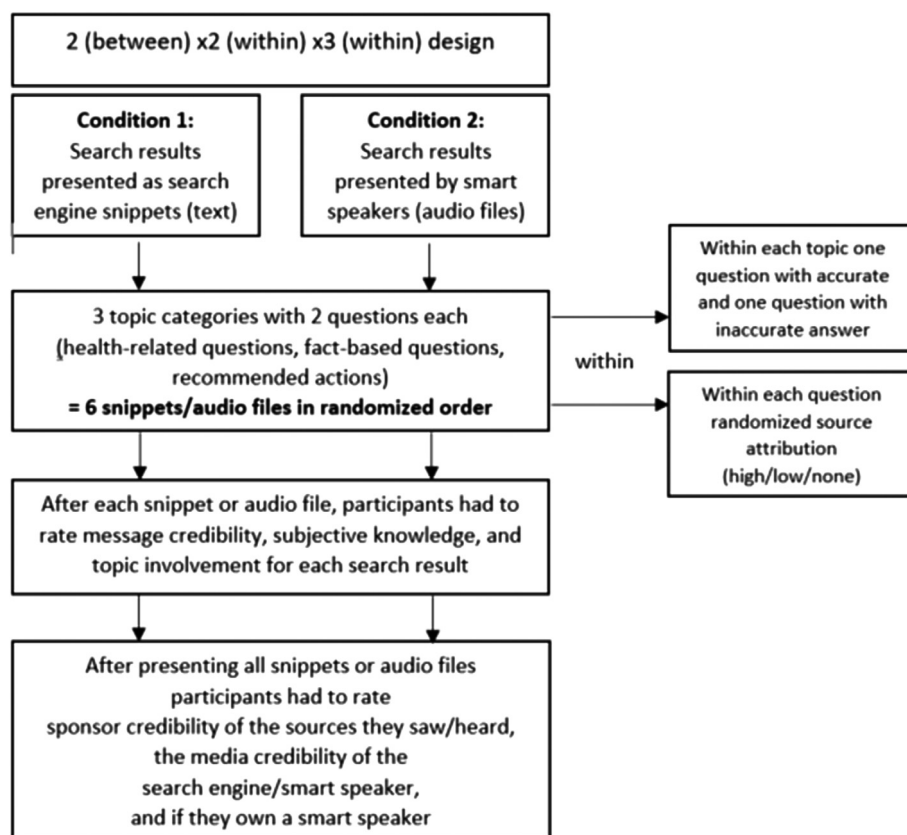
Figure 1. Overview of procedure.

content-related factors play and, more importantly, if and how they interact with modality. On the basis of the MAIN model (Sundar, 2008) and contradictory findings about different evaluations in textual and auditive modalities during information processing, we examine whether message credibility perceptions differ between a text-based (search engine) and voice-based (smart speaker) message modality. Additionally, we build on work on sponsor credibility (Metzger et al., 2010; Westerwick, 2013) and the accuracy of content (Tate, 2018) to explore in which modality users are more sensitive to these credibility cues.

A 2 (modality: text/voice) × 2 (message accuracy: high/low) × 3 (sponsor credibility: high/low/none) mixed factorial experimental design was employed online (procedure can be seen in Figure 1). We used a two-study approach, where we first exploratively asked research questions for some effects and aimed to replicate the found effects in a second experiment.

Theoretical Background

Credibility on the Web
People rely heavily on the Internet when searching for information. However, due to the amount of information, there is also the danger of inaccurate or biased information

(Metzger et al., 2003), and people might wonder whether the information they found online is correct (Jung et al., 2016). One criterion people use to filter information is credibility. Credibility is based on two key dimensions: trustworthiness and expertise (Lewandowski, 2012).

Models on credibility assessments suggest that users can draw on different credibility cues when considering the source, website, and message attributes to make credibility judgments (Flanagin & Metzger, 2007; Sundar, 2008). Therefore, Flanagin and Metzger (2007) argue to differentiate between types of credibility. We will focus on message credibility as the primary outcome variable in this study, which is the "individual's judgment of the veracity of the content of communication" (Appelman & Sundar, 2016, p. 63).

Technological Influences on Message Credibility

Modality as a Credibility Cue
People can get information from all kinds of media incorporating different technological affordances. These affordances can be relevant for credibility judgments. In his MAIN model, Sundar (2008) proposes that "technological affordances in digital media trigger cognitive heuristics that

aid credibility judgments [...]" (p. 78). He identified different affordances of technologies that can trigger heuristic cues that may play a role in credibility assessment. Modality, that is, whether information is delivered textually, aurally, or audiovisual, is the most structural and apparent affordance when using an interface (Sundar, 2008). According to the model, each modality can, by its sheer presence, cue heuristics that can affect credibility judgments. Smart speakers without a screen operate on only one modality: auditive.

Sundar (2008) identified different heuristics that can be triggered by modality. According to the "realism heuristic" (Reeves & Nass, 1996; Sundar, 2008), meaning modalities that are promoting realism, modalities conveying a more human-like interaction are likely to increase trust as their content resembles the real world (Sundar, 2008, p. 80). Thereby, audio has been said to be important for promoting more human-like interaction (Sundar, 2008). Smart speakers might even have more potential because interacting includes hearing and also spoken dialogue. The voice gives the assistants a human-like appeal, making the interaction seem even more human-like (Cho et al., 2019). A concept that intertwines with this is anthropomorphism. In the context of agents, "anthropomorphism describes the phenomenon of attributing human-like traits to these technologies" (Li & Suh, 2021, p. 4054). Anthropomorphism can influence the adoption (Moussawi et al., 2021), perception, and usage of voice assistants (Wagner & Schramm-Klein, 2019). Additionally, Qiu and Benbasat (2009) found that anthropomorphized voices lead to higher credibility evaluations due to the feeling of social presence when using recommendation agents. Therefore, the output of the smart speaker, that is, operating with a more human-like voice, could trigger anthropomorphism and positively affect credibility.

Another factor that could be of relevance is cognitive load. Research in human–computer interaction and linguistics showed that auditive and textual modalities differ in terms of "transmission and reception" (Rzepka et al., 2022). Getting information in different modalities can affect the amount of cognitive load people are experiencing during information processing and credibility perceptions (Tabbers et al., 2004). Research on cognitive load theory (Sweller, 1999) and multimedia learning (Mayer, 2001) showed that changing the presentation format can alter the cognitive load a person is experiencing. For example, presenting information aurally instead of visually in learning settings can lead to lower mental effort and improve learning processes (Kalyuga et al., 2000). However, this line of research focuses on instructions and learning processes and not on everyday common information search and credibility as an outcome. By contrast, Ischen et al. (2021) found that communicating with a text-based virtual assistant led to less cognitive processing compared to communicating with a voice-activated one.

Studies concerning modality effects and credibility perceptions have revealed mixed results. Some studies found that additional visual and auditive stimuli lead to higher credibility ratings when comparing television news with newspapers (Ibelema & Powell, 2001; Metzger et al., 2003). In a recent study on messaging apps, Sundar et al. (2021) found that video content was perceived as more credible than auditive or textual content due to the realism heuristic. However, there was no significant difference between audio and text. Other studies found that textual modalities led to more positive evaluations from receivers than auditive or audiovisual modalities, for example, for online news (Kiousis, 2001; Sundar, 2000). Concerning voice-activated assistants, Cho et al. (2019) found that voice interaction leads to more positive attitudes toward the assistants than text.

Summarizing, research about the effects of different message modalities is inconclusive and mainly focused on online news. Additionally, the few existing studies including voice assistants did not look at credibility perceptions as a possible outcome. Due to the inconsistent results, we pose an open research question on differences between text-based output with search engines and voice-based output with smart speakers.

Research Question 1 (RQ1): Is there a difference in the assessment of perceived message credibility when receiving information in a text-based (search engine) or voice-based (smart speaker) modality?

Content-Related Influences on Message Credibility

Sponsor Credibility

Next to modality, content-related factors can affect credibility judgments. Studies on web credibility have already demonstrated that the source of information is crucial. For example, Metzger et al. (2010) found that one of the most prevalent heuristics that users rely on is the source. Information that credits known or established organizations or websites is perceived as more credible. Therefore, source cues related to expertise can play an important role in individuals' perception of credibility (Hu & Sundar, 2010). According to Sundar (2008), this is one of the most robust findings in credibility research. Unkel and Haas (2017) could demonstrate this in the context of search engine results.

In the context of web search, sponsor credibility, that is, the "evaluation of the website's sponsor, which may result from expertise or personal experience with the organization, group, or person" (Westerwick, 2013, p. 196), is an

important source cue that positively affects message credibility (Westerwick, 2013). We adapted this term to see if this heuristic is present when using a voice-based technology and modality. This is important because users get to hear only one source or none during voice search. Therefore, we compare a high-reputation with a low-reputation source condition and a no-source condition and propose that higher sponsor credibility will lead to higher message credibility ratings across modalities. More importantly, we explore whether these sponsor effects differ between the two modality conditions.

Hypothesis 1 (H1): Higher (vs. low/none) sponsor credibility leads to higher perceived message credibility of the information in both modalities.

Research Question 2 (RQ2): Does sponsor credibility have a stronger effect in one of the modality conditions?

Message Accuracy

Another major factor for credibility judgments of online information is the quality. The most central dimension for determining information quality is message accuracy (Fogg et al., 2001; Rieh & Belkin, 1998; Tate, 2018). Accuracy can be defined as "the extent to which information is reliable and free of errors" (Kammerer & Gerjets, 2012, p. 254). Reinhard and Sporer (2010) mention that the level of message accuracy comes into play when recipients form credibility perceptions. In line with this, Metzger et al. (2003) found that inaccurate news stories decrease credibility perceptions. Additionally, Jung et al. (2016) found in a study about credibility of online information that higher message accuracy led to increased credibility perceptions.

We expect to replicate this well-established effect also for a smart speaker-based online search. Since people get only one answer and cannot quickly look at other results or sources, message accuracy seems even more important. Therefore, we explore in which output modality users are more sensitive to the accuracy of the information. According to the distraction heuristic, a new or more multimodal experience – like talking and listening to a machine – could lead to sensory overstimulation, possibly distracting users from effortfully evaluating the content (Sundar, 2008). Again, this could be connected to higher cognitive load during auditive information processing, leading to less critical evaluation (Sweller, 1999). For example, receiving video compared to text-based information can decrease processing depth (Powell et al., 2018), which could be especially important for accurate evaluations. Also, Sundar et al. (2021) find proof that people process audiovisual material (videos) more superficially than text-based information in the context of WhatsApp messages and are more likely to believe false information. Therefore, message accuracy

might be assessed and evaluated to different degrees in different modalities. However, Sundar et al. (2021) did not find a difference between voice and text in the messenger context. Since we look at this in the context of smart speaker and search engine output, we propose a research question to investigate a possible interaction between message accuracy and modality.

Hypothesis 2 (H2): Higher message accuracy (accurate information) leads to higher perceived message credibility of the information in both modalities.

Research Question 3 (RQ3): Does message accuracy have a stronger effect in one of the modality conditions?

Topic Involvement

Lastly, audience factors such as topic involvement can influence users' perceptions of information. Topic involvement can be defined as the degree to which users find an issue relevant or how much they care about a specific topic (Jung et al., 2016; Petty & Cacioppo, 1990; Westerwick, 2013). Based on the elaboration likelihood model of persuasion (ELM; Petty & Cacioppo, 1990), the level of topic involvement plays a role in information processing. Studies show that a higher level of involvement can lead to more systematic processing of the message, whereas a lower level results in more heuristic processing (Chaiken, 1980; Petty & Cacioppo, 1990; Reinhard & Sporer, 2010). Concerning sponsor credibility and message accuracy, this means that content factors such as accuracy are more influential when involvement is high because the motivation to critically scrutinize the message is higher (Westerwick, 2013), and sponsor credibility, as a heuristic cue, becomes more important when topic involvement is low. According to Petty et al. (1981), argument quality increases persuasion under high-involvement conditions, and source credibility influences persuasion under low-involvement conditions (Jung et al., 2016).

Therefore, we presume that topic involvement also plays a role during information search in both modalities and propose:

Hypothesis 3 (H3): Low topic involvement will positively moderate the effects of sponsor credibility on perceived credibility in both modalities, that is, for people with low involvement the effect of high sponsor credibility on message credibility will be stronger.

Hypothesis 4 (H4): High topic involvement will positively moderate the effects of message accuracy on perceived credibility in both modalities, that is, for people with high involvement the effect of high message accuracy on message credibility will be stronger.

Method

Design and Sample

We employed a 2 (modality: text/voice; between) × 2 (message accuracy: high/low; within) × 3 (sponsor credibility: high/low/none; within) mixed factorial experimental design with 399 participants. Participants were recruited via the platform prolific and paid for their participation. They had to live in Germany and be fluent in German. Of the 399 participants, 60% were male, 39% female, and 1% other. On average, participants were 30.3 (SD = 9.9) years old. Age and gender did not significantly differ between the two experimental groups. The study was pre-registered (https://aspredicted.org/2y8er.pdf).

Procedure

Participants were randomly assigned to the two modality conditions and told they would see search results of either a new search engine or smart speaker called *Yoursearch*. Participants received six search results from three topic categories that varied in message accuracy (high/low) and sponsor credibility (high/low/none). In the text output condition, these were presented in the style of *search engine snippets* to make the two modalities comparable (only one result). In the voice output condition, participants received *audio files* containing identical text but read by a smart speaker. After each stimulus, participants had to rate the message credibility of the content, their topic involvement, and subjective knowledge about the question. After seeing all stimuli, participants had to rate the media credibility of the search engine/smart speaker and the sponsor credibility of all the sources they saw. An attention check was included to control whether people were listening to the audio files in the voice condition. An overview of the experimental procedure can be found in Figure 1.

Stimuli

We chose three topic categories to cover different types of questions to control for potential topic effects. The first category consists of two health-related questions about the health dangers of, for example, e-cigarettes. Second, we chose two fact-based questions, such as the number of bones in a human body, to include more mundane scenarios since many people use voice-based search for factual rather than more complex questions. Lastly, we included two questions about recommended actions, such as how to act when meeting a wolf, because this could have more direct consequences. In the end, 12 text snippets and voice files were created. One answer was completely accurate in every topic category, and the other contained false

information. For example, while the answer to a question concerning stopping nosebleeding gave correct advice, a recommendation regarding meeting a wolf gave wrong advice, for instance, feeding it. Additionally, for each participant, every question randomly credited a highly credible, a low-credible, or no source. High-credible sources included websites of official authorities or public-law broadcasters. For low-credible sources, we used commercial websites, lifestyle magazines, or Q&A portals. The voice files were created with the Skill Blueprint function from Amazon Alexa. In the end, participants were debriefed about the aim of the study and the material. In the first experiment, we had an additional but less central RQ on recall, which we excluded from this manuscript due to less thematic relevance and due to the replication study. The material, the dataset, and the analysis script are available at https://osf.io/bqtzm/.

Measures

Message Credibility
We assessed message credibility with a scale by Appelman and Sundar (2016). Respondents had to indicate how well the adjectives "authentic," "believable," and "accurate" describe the information they just read/heard on scales ranging from 1 = *describes it very badly* to 7 = *describes it very well* (α = .91). This variable and all the following were averaged over all their respective items (Exp1: M = 5.55, SD = 1.22, Min. = 1.0, Max. = 7.0; Exp2: M = 5.65, SD = 1.26, Min. = 1.0, Max. = 7.0).

Topic Involvement
To measure involvement, we used a 4-item scale by Flanagin and Metzger (2007). Participants had to agree to statements such as, "The information was very relevant for my own life" on scales ranging from 1 = *does not apply at all* to 7 = *fully applies* (α = .89; Exp1: M = 4.51, SD = 1.3, Min. = 1.0, Max. = 7.0; Exp2: M = 4.50, SD = 1.29, Min. = 1.0, Max. = 7.0).

Subjective Previous Knowledge
Here, we asked participants: "Could you have answered the question before reading/hearing the answer?" on a scale from 1 = *not at all* to 7 = *completely* (Exp1: M = 3.96, SD = 1.74, Min. = 1.0, Max. = 7.0; Exp2: M = 4.07, SD = 1.77, Min. = 1.0, Max. = 7.0)

Sponsor Credibility
We measured sponsor credibility with a 5-item scale by Flanagin and Metzger (2007), including items such as "[Website Name] has a very good reputation of posting valuable information." Participants had to agree on scales from 1 = *not at all* to 7 = *completely* (α = .95; Exp1: M = 4.23, SD = 1.01, Min. = 1.0, Max. = 6.9; Exp2: M = 4.25, SD = 1.03, Min. = 1.0, Max. = 6.9).

Media Credibility

To measure credibility perceptions of the smart speaker/ search engine, we used a scale by Flanagin and Metzger (2000). Respondents were asked about their perception concerning the five items "believable," "accurate," "trustworthy," "biased," and "complete" on scales ranging from 1 = *not at all* to 7 = *extremely* (α = .83; Exp1: $M = 5.05$, $SD = 1.06$, Min. = 1.0, Max. = 7.0; Exp 2: $M = 4.93$, $SD = 0.98$, Min. = 1.6, Max. = 7.0)

Demographics

We assessed gender, age, and education level (see OSF for details).

Results

Manipulation Checks

Using a linear mixed effect model with manipulated sponsor credibility as the predictor and perceived sponsor credibility as the dependent measure, we saw a significant impact of the manipulation, $\chi^2(2) = 2254.83$, $p < .001$. A post hoc test showed that ratings were higher for sponsors with higher manipulated sponsor credibility ($M = 5.23$, $SD = 1.19$) than for sponsors with low credibility, $M = 3.20$, $SD = 1.21$, $t(1,602) = -33.98$, $p < .001$, $d = 1.70$, indicating a successful manipulation.

When using the item "accurate" from the message credibility scale as the dependent measure, manipulated accuracy had a significant effect, $\chi^2(1) = 108.78$, $p < .001$. Participants rated the highly accurate messages as more accurate ($M = 5.73$, $SD = 1.14$) than the low-accuracy messages, $M = 5.19$, $SD = 1.56$, $t(1,196) = 11.62$, $p < .001$, $d = .34$. The participants in the two groups did not differ in previous knowledge – search engine (SE): $M = 3.93$, $SD = 1.73$; smart speaker (SM): $M = 4.00$, $SD = 1.75$, $t(2,391) = .97$, $p = .33$.

Research Questions and Hypotheses

To answer RQ1, we conducted a *t* test comparing message credibility ratings in the two modality conditions. Participants in the smart speaker modality perceived the same information as significantly more credible ($M = 5.66$, $SD = 1.11$) than did people in the search engine condition, $M = 5.44$, $SD = 1.31$, $t(383) = -2.63$, $p < .01$, $d = .26$. Additionally, we exploratively found that participants also rated the smart speaker as overall more credible ($M = 5.21$, $SD = 0.97$) compared to the search engine in terms of media credibility, $M = 4.89$, $SD = 1.12$; $t(388) = -3.08$, $p < .01$, $d = .31$.

We conducted a linear mixed-effect model to test H1 and RQ2. As shown in Table 1 (created with sjPlot; Lüdecke, 2021), the model revealed a significant main effect of the manipulated sponsor credibility on perceived message credibility, $\chi^2(2) = 39.88$, $p < .001$. More importantly, there was a significant interaction between sponsor credibility and modality, $\chi^2(2) = 15.31$, $p < .001$. A low-credible sponsor decreased message credibility attributions in both modalities, which is in line with H1 – high sponsor credibility: (SE) $M = 5.65$, $SD = 1.08$, (SM) $M = 5.80$, $SD = 1.03$; low sponsor credibility: (SE) $M = 5.30$, $SD = 1.38$, (SM): $M = 5.40$, $SD = 1.27$. This was not the case for no-sponsor attribution. Not adding a sponsor decreased message credibility perception in the text-based condition as much as a low-credible sponsor (no sponsor: $M = 5.37$, $SD = 1.43$), but not in the voice-based condition. Here, message credibility ratings in the no-sponsor condition were as high as in the high-credible sponsor condition (no sponsor: $M = 5.79$, $SD = 0.97$). A post hoc test showed significantly higher message credibility ratings for information with no sponsor in the auditive modality than in the textual modality, $t(691) = -4.82$, $p < .001$, $d = .34$).

Concerning H2 and RQ3, the linear mixed-effect model showed a significant main effect of message accuracy on perceived message credibility, $\chi^2(1) = 120.55$, $p < .001$ (Table 1). In both modality conditions, accurate information produced higher message credibility ratings than the low-accuracy information, supporting H2. Additionally, we found a significant interaction between message accuracy and modality, $\chi^2(1) = 6.90$, $p < .01$. The reduction in message credibility ratings when the accuracy of the content was low was larger in the textual modality – high accuracy: $M = 5.79$, $SD = 0.97$; low accuracy: $M = 5.16$, $SD = 1.46$ – than in the auditive modality – high accuracy: $M = 5.79$, $SD = 0.97$; low accuracy: $M = 5.48$, $SD = 1.25$. A post hoc test showed significantly higher message credibility ratings for information of low message accuracy in the auditive modality than in the text-based condition, $t(1,165) = -4.13$, $p < .001$, $d = .24$.

Lastly, we looked at the moderating effect of topic involvement. The model revealed a significant interaction between topic involvement and sponsor credibility, $\chi^2(2) = 16.08$, $p < .001$. When topic involvement was high, message credibility ratings were generally high regardless of the sponsor conditions (see Figure 2). In line with H3, the effect of a highly (vs. low) credible sponsor on message credibility ratings was significantly larger when topic involvement was low in both modalities. There was an unpredicted significant three-way interaction between topic involvement, sponsor credibility, and modality, which is important to consider when discussing the two-way interaction, $\chi^2(2) = 17.19$, $p < .001$. The pattern of the two-way interaction between sponsor credibility and modality was stronger when topic involvement was low versus high (model table can be found on OSF). In Figure 2, we see that the lines for no and high sponsors have almost no distance in the

Table 1. Linear mixed models for sponsor credibility (H1; RQ2) and message accuracy (H2; RQ3) – Experiment 1

Predictors	Message credibility (H1; RQ2)				Message credibility (H2; RQ3)			
	Estimates	CI	p	df	Estimates	CI	p	df
(Intercept)	5.65	5.50 – 5.79	**<.001**	750.57	5.73	5.60 – 5.86	**<.001**	565.08
Modality [voice]	0.14	−0.06 – 0.35	.164	772.26	0.12	−0.07 – 0.30	.208	565.08
Sponsor [low]	−0.30	−0.45 – −0.15	**<.001**	659.07				
Sponsor [no]	−0.31	−0.46 – −0.17	**<.001**	656.08				
Modality [voice] × Sponsor [low]	−0.08	−0.29 – 0.12	.426	648.15				
Modality [voice] × Sponsor [no]	0.31	0.10 – 0.52	**.003**	654.24				
Accuracy [low]					−0.57	−0.68 – −0.46	**<.001**	397.00
Modality [voice] × Accuracy [low]					0.21	0.05 – 0.36	**.009**	397.00
Random effects								
σ^2	0.90				0.86			
τ_{00}	0.04 $_{ID:sponsor}$				0.02 $_{ID:accuracy}$			
ICC	0.52 $_{ID}$				0.54 $_{ID}$			
	0.39				0.40			
N	399 $_{ID}$				399 $_{ID}$			
	3 $_{sponsor}$				2 $_{accuracy}$			
Observations	2,394				2,394			
Marginal R^2/Conditional R^2	0.026/0.401				0.047/0.424			

Note. The marginal R^2 considers only the variance of the fixed effects, while the conditional R^2 takes both the fixed and random effects into account. The computation of p values is based on conditional F tests with Kenward–Roger approximation for the degrees of freedom. CI = 95% confidence intervals (lower and upper bound); df = degrees of freedom; ICC = intraclass correlation coefficient. Bold p values indicate significant results.

auditive condition. In the textual condition, we see a greater distance when topic involvement is low.

Concerning message accuracy and topic involvement (H4), we found a significant two-way interaction, $\chi^2(1) = 23.47$, $p < .001$. The effect of message accuracy on message credibility ratings was larger when topic involvement was lower and not when involvement was higher than predicted (see Figure 3). Thus, H4 is not supported (model table can be found on OSF).

Discussion

This experiment provided first insights into how modality affects the perceived credibility of search results. Regarding RQ1, we found that people perceive information conveyed by a smart speaker as more credible than the same information by a text-based search engine. Moreover, concerning H1 and RQ2, we found that sponsor credibility plays a role. High-credible sponsors led to higher message credibility ratings in both modalities than did low-credible sponsors. But in the search engine condition only, this was also the case compared to no-source attribution. Giving no source did not decrease message credibility ratings in the smart speaker modality, which indicates that knowing the source might be less important for people's credibility judgments during voice search.

In line with prior research, we furthermore showed that message accuracy affects message credibility perceptions.

In both modalities, higher accuracy led to higher message credibility ratings (which supports H2). Interestingly, concerning RQ3, the analysis showed that this effect was stronger in the search engine condition than in the smart speaker condition, indicating that people are less sensitive to incorrect information when listening to smart speakers.

Lastly, we could show that the level of involvement plays a role. The effects of sponsor credibility and message accuracy on message credibility were moderated by topic involvement in both cases, such that the effects were larger when topic involvement was low. This supports H3 but not H4. Interestingly, we found an unpredicted three-way interaction for sponsor credibility, involvement, and modality. It showed that the differences between the modalities concerning the effect of no-sponsor attribution are larger when people are less involved in the topic. To demonstrate the replicability of these findings, we conducted a preregistered replication of the findings in a second experiment.

Experiment 2

Building on the results of the first experiment, we proposed the following hypotheses for the replication:

Hypothesis 1 (H1): Participants in the smart speaker modality will perceive the given information to be more credible compared with people in the search engine modality.
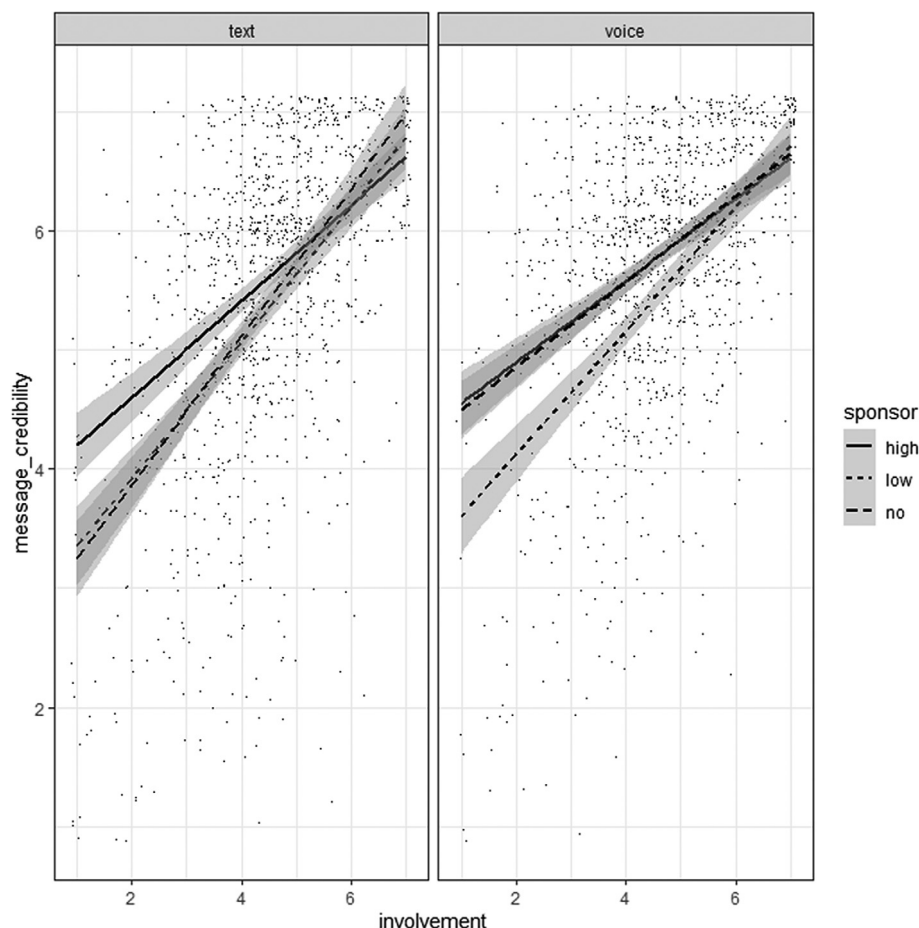
Figure 2. Experiment 1: Interaction plot: modality × sponsor credibility × involvement.

Hypothesis 2 (H2): Participants in the smart speaker modality will rate the presented smart speaker as more credible compared with how people in the search engine modality will rate the presented search engine.

Hypothesis 3 (H3): There will be an interaction effect between modality (smart speaker) and message accuracy (low) consistent with the findings from Experiment 1.

Hypothesis 4 (H4): There will be an interaction effect between modality (smart speaker) and sponsor credibility (no-sponsor condition), consistent with the findings from Experiment 1.

Hypothesis 5 (H5): There will be a three-way interaction between modality (smart speaker), sponsor credibility (no-sponsor condition), and topic involvement, consistent with the findings from Experiment 1.

Design and Sample

To replicate the findings, we used the same experimental design and recruitment strategy as in Experiment 1. Of the 398 participants, 58% were male, 41% female, and 1% other. On average, participants were 28.2 ($SD = 8.34$) years old. Age and gender did not significantly differ between the two experimental groups. The replication was also preregistered (https://aspredicted.org/vb4u2.pdf).

Stimuli

The stimuli and measures were adopted from the first experiment.

Results

Manipulation Checks

Using the same linear mixed-effect model as in Experiment 1, we saw a significant impact of manipulated sponsor credibility, $\chi^2(2) = 2{,}210.32$, $p < .001$. Highly credible sponsors led to higher sponsor credibility ratings ($M = 5.121$, $SD = 1.18$) than did low-credible sponsors ($M = 3.18$, $SD = 1.22$), meaning the manipulation worked as intended, $t(1{,}321) = -29.44$, $p < .001$, $d = 1.62$.

The manipulation check for message accuracy was also successful, $\chi^2(1) = 66.43$, $p < .001$. Participants rated the
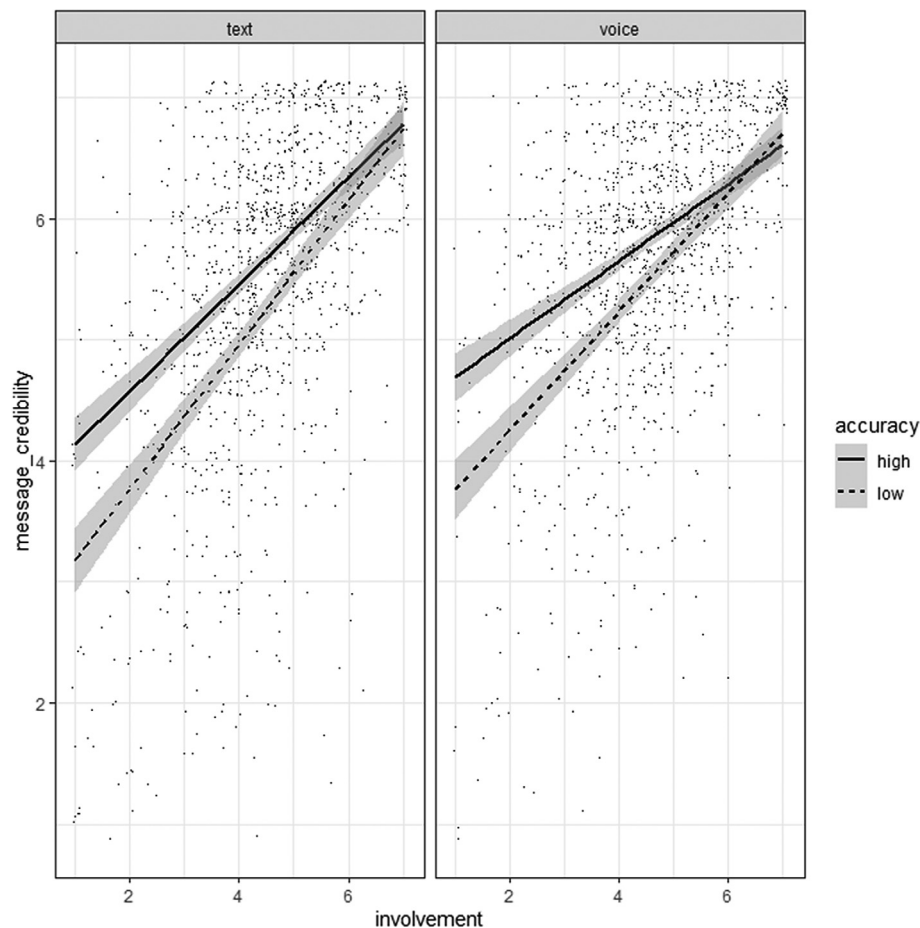
Figure 3. Experiment 1: Interaction plot: modality × message accuracy × involvement.

messages with high accuracy as more accurate (M = 5.70, SD = 1.20) than the messages with low accuracy (M = 5.35, SD = 1.46), t(1,193) = 8.99, p < .001, d = .26. Again, the participants in the two groups did not differ in previous knowledge – SE: M = 4.08, SD = 1.72; SM: M = 4.06, SD = 1.81; t(2,380) = −.37, p = .71.

Hypotheses Tests
To answer H1 and H2, we conducted two t tests. For H1, we compared message credibility ratings in the two modality conditions. Participants in the smart speaker modality again perceived the same information as significantly more credible (M = 5.75, SD = 1.18) than did people in the search engine condition (M = 5.54, SD = 1.32), t(396) = −2.50, p < .05, d = .25, which supports H1. We also compared overall media credibility ratings, t(389) = −1.87, p =.06. The smart speaker did not receive significantly better media credibility ratings (M = 5.02, SD = 0.91) than the search engine (M = 4.84, SD = 1.04), which does not support H2.

To investigate H3, we rerun the linear mixed-effect model (Table 2). We replicated the interaction between message accuracy and modality on perceived message credibility, χ^2(1) = 6.36, p < .05. Again, the reduction in

message credibility ratings was larger in the textual modality (high accuracy: M = 5.84, SD = 1.00; low accuracy: M = 5.24, SD = 1.52) than in the auditive modality (high accuracy: M = 5.94, SD = 1.03; low accuracy: M = 5.56, SD = 1.29). Mean comparisons using a post hoc test showed significantly higher message credibility ratings for information of low message accuracy in the auditive modality than in the text-based condition, t(1162) = −3.90, p < .001, d = .23.

In contrast to H4, we did not find a significant interaction of manipulated sponsor and modality on perceived credibility, χ^2(2) = 3.52, p = .17. We saw a different pattern in the search engine modality than in Experiment 1. No-sponsor attribution did not lead to a decrease in message credibility in the search engine modality compared to a high-credible sponsor (high sponsor credibility: M = 5.60, SD = 1.28; no sponsor: M = 5.58, SD = 1.31). More importantly, we found the same pattern in the smart speaker modality as in Experiment 1. Not adding a sponsor again did not decrease message credibility perception in the smart speaker condition compared to a highly credible sponsor (high sponsor credibility: M = 5.81, SD = 1.67; no sponsor credibility: M = 5.91, SD = 1.11). Participants reported higher message credibility ratings for information with no sponsor in the auditive
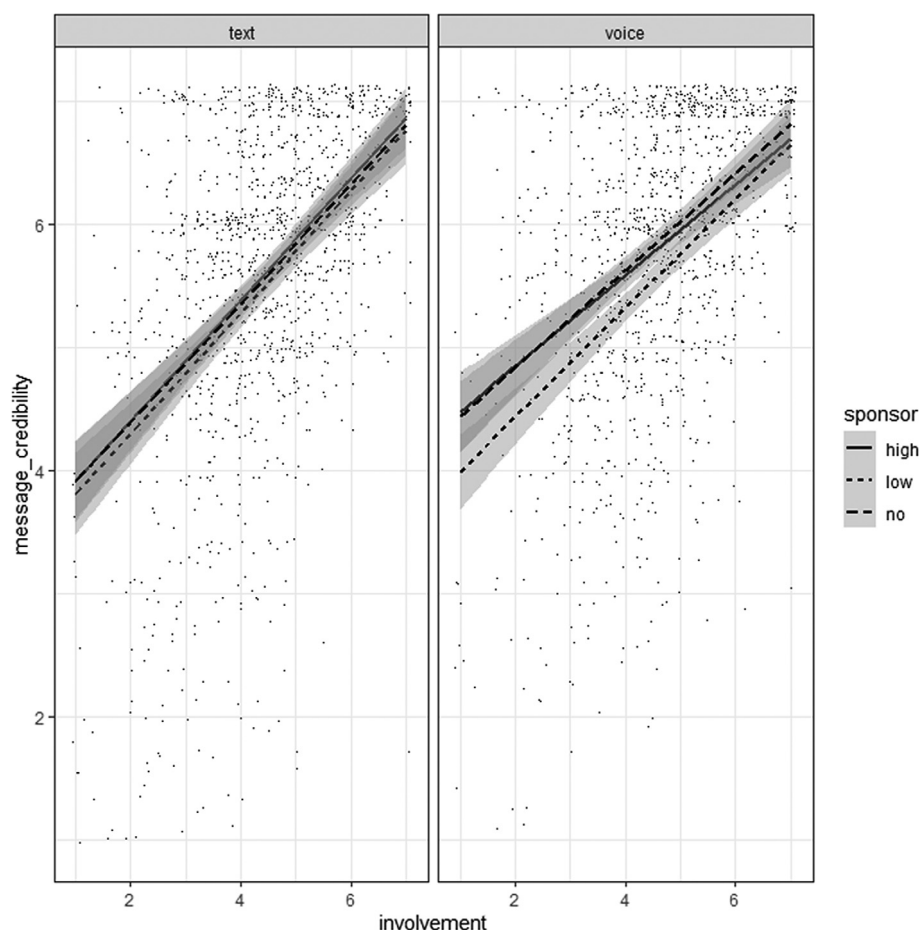
Figure 4. Experiment 2: Interaction plot: sponsor credibility × modality ×involvement.

modality compared with respondents in the text-based condition, $t(776) = -3.75$, $p < .001$, $d = .27$.

Lastly, concerning H5, the model did not show a significant three-way interaction between topic involvement, sponsor credibility, and modality, $\chi^2(2) = 0.74$, $p = .69$. However, when topic involvement was low, we saw the same pattern in the smart speaker modality as in Experiment 1 but not in the search engine modality (see Figure 4, model table can be found on OSF).

General Discussion

This two-study paper explored how message modality can impact credibility judgments during information search and what role content-related factors play. Findings showed that information presented by smart speakers is perceived as more credible, and message credibility perceptions are less affected by the lack of source information and low message accuracy in the auditive modality. These findings not only replicate previous effects on web credibility in the context of smart speakers but also offer new insights by

showing that people evaluate information differently when getting information aurally. The results point to potential downsides of smart speaker usage for information search.

In both experiments, we found that information presented by smart speakers was rated as more credible than information from classic search engines, which supports the prediction of the MAIN model (Sundar, 2008) that if the same information is delivered aurally, it can be perceived as more credible. These findings are in line with previous research that found content from auditive modalities to be rated as more credible than content from textual modalities (Ibelema & Powell, 2001; Metzger et al., 2003). We extended this to a relatively novel technology: smart speakers. Voice interaction can, thus, lead to not only a more positive attitude toward the assistant (Cho et al., 2019) but also to higher credibility perceptions of the content it provides. This could indicate that heuristics positively affecting credibility perceptions, such as the realism heuristic (Sundar, 2008) or anthropomorphism conveyed through the voice output (Qiu & Benbasat, 2009), are more likely to be applied when using smart speakers. Next, future

Table 2. Linear mixed models for message accuracy (H3) and sponsor credibility (H4) – Experiment 2

Predictors	Message credibility (H3)				Message credibility (H4)			
	Estimates	CI	p	df	Estimates	CI	p	df
(Intercept)	5.75	5.63 – 5.88	**< .001**	557.51	5.71	5.57 – 5.85	**< .001**	710.13
Modality [voice]	0.11	−0.06 – 0.29	.211	557.51	0.12	−0.07 – 0.32	.210	711.39
Accuracy [low]	−0.50	−0.60 – −0.39	**< .001**	397.00				
Modality [voice] × Accuracy [low]	0.19	0.04 – 0.34	.013	397.00				
Sponsor [low]					−0.31	−0.45 – −0.18	**< .001**	601.84
Sponsor [no]					−0.19	−0.34 – −0.05	**.007**	656.29
Modality [voice] × Sponsor [low]					−0.05	−0.24 – 0.14	**.584**	589.38
Modality [voice] × Sponsor [no]					0.28	0.08 – 0.48	**.006**	657.15
Random effects								
σ^2	0.78				0.80			
	$0.03_{ID:accuracy}$				$0.01_{ID:sponsor}$			
τ_{00}	0.53_{ID}				0.53_{ID}			
ICC	0.41				0.40			
N	399_{ID}				399_{ID}			
	$2_{accuracy}$				$3_{sponsor}$			
Observations	2,394				2,263			
Marginal R^2/Conditional R^2	0.038/0.439				0.027/0.416			

Note. The marginal R^2 considers only the variance of the fixed effects, while the conditional R^2 takes both the fixed and random effects into account. The computation of p values is based on conditional F tests with Kenward–Roger approximation for the degrees of freedom. CI = 95% confidence intervals (lower and upper bound); df = degrees of freedom; ICC = intraclass correlation coefficient. Bold p values indicate significant results.

research should study which processes are responsible for these effects in detail. Nevertheless, our findings imply that quality standards of voice query results should be high since users are more likely to find them credible. This could be challenging concerning the one-result answer technique of smart speakers.

Concerning content-related factors, we were able to replicate prior findings on source reputation and extend these to smart speaker output (Hu & Sundar, 2010; Metzger et al., 2010). Information that credits sponsors with high reputations is perceived as more credible than low-credible sponsors in both modalities. But interestingly, we found differences between the two modalities when no sponsor was given. Even though we could not replicate the two-way interaction between sponsor credibility and modality, we still found the same pattern for the voice output modality in both experiments. Providing no sponsor information results in the same credibility ratings as naming a highly credible sponsor. That no-sponsor attribution does not negatively affect credibility ratings is important because smart speakers often do not give the provenance of information (Dambanemuya & Diakopoulos, 2021). Consequently, people could perceive information from less credible sources to be as credible as information from highly credible sources without a source attribution, which could be problematic for adequate credibility judgments. These findings highlight the importance of sponsor credibility when the medium operates on an auditive modality.

Additionally, message accuracy affected credibility in both modalities. Higher message accuracy led to higher message credibility ratings, which supports previous findings (Jung et al., 2016; Metzger et al., 2003). However, the more important finding is that with voice output, users seem less sensitive to accuracy because incorrect information did not lead to the same decrease in credibility ratings as with search engines. Since accuracy is one of the most important factors for evaluating information quality (Tate, 2018), people should spot inaccurate information. Our results indicate that less critical evaluation happens when information is heard instead of read. People might experience more cognitive load when processing information aurally compared to text.

That users are less likely to question information from smart speakers is especially problematic when incorrect or biased information is presented, as discovered by *The New York Times* (Thompson, 2022). Even if, currently, questions tend to be more basic, this could become highly relevant when people start to use speakers for more complex or controversial topics.

Lastly, it is important to mention the role of topic involvement. Consistent with the ELM (Petty & Cacioppo, 1990; Westerwick, 2013), higher sponsor credibility as a heuristic cue became more important when involvement was low. More importantly, we have the first indications that the effect concerning no-sponsor attribution in the auditive modality might be stronger when topic involvement is low (albeit we could not replicate the significant three-way interaction from Experiment 1).

Against our expectations and predictions from the ELM (Petty & Cacioppo, 1990), we found the same pattern for

message accuracy. The difference in credibility ratings for high- and low-accuracy information was larger when topic involvement was low. One explanation could be that people with low topic involvement spent more time processing since it was new information, and they wanted to perform well in the study. This could be different in a real situation. It could also be that people with higher involvement only skimmed the information since they thought they already knew it.

Considering all findings, the stakes for the quality of the single response of smart speakers are even higher than expected (Dambanemuya & Diakopoulos, 2021). This can have multiple implications for users and the companies behind the devices. Since users seem to give the benefit of the doubt to voice answers, they should be made more aware of the fact that a single speaker answer is not automatically correct. Additionally, users' source awareness should be strengthened by recommending they pay attention to source attribution and verify the information when in doubt.

On the company side, the results could be considered in future designs of voice search interactions. To help users make adequate credibility evaluations, smart speaker results could always mention the source or present multiple answers to choose from. It might also be helpful to highlight whether there are contradictory answers before giving a result. We saw that people consider source information when receiving it. This indicates that people generally learned to watch for these source cues, but with voice output, they do not notice it much when those cues are missing, making it even more important to give a source.

Limitations

Naturally, this study has several limitations. First, participants did not complete the task in an authentic scenario. Especially for the voice-based condition, it must be considered that we used pre-recorded audios instead of actual interactions. In everyday life, people often do other activities alongside asking their smart speaker questions, so there might be even less critical evaluation because people's main attention is not solely on the answer as it was in the experiment. Further, some of the effects we found are rather small, which should be taken into consideration. However, these effects were replicated in the second experiment, which speaks against negligible effects. Additionally, we used Amazon Alexa as the voice. Some participants might be familiar with the sound and have certain expectations due to previous experience. Future studies should try to apply more authentic scenarios closer to actual usage situations.

The goal of the present study was to investigate whether modality affects credibility perceptions during information search and which factors influence this effect. As already mentioned, the next step would be to investigate the underlying processes. Future research could investigate which of the various heuristics people apply, for example, human-like cues of the smart speaker's voice. Additionally, one could measure the cognitive load people are experiencing during voice search to see if this leads to different information processing. This could be especially interesting since, in real life, people also do other activities while communicating with their smart speaker.

Conclusion

With this two-study approach, we investigated how message modality as a technological factor of smart speakers can affect credibility perceptions. Through this, we were able to gain important insights. We showed that information presented by smart speakers (auditive modality) was perceived as more credible than the same information from search engines (textual modality). Additionally, we found that the absence of source information did not lower credibility perceptions in the voice-output condition, which is important since smart speakers often do not provide any source attribution. Moreover, users seem to be less able or motivated to identify incorrect information in voice than in text output. All of these findings are highly relevant when thinking about how essential it is for users to be able to judge the credibility of information adequately. Therefore, these first insights should be considered in the future when evaluating voice search and for prospective research and design of voice search interactions.

References

Appelman, A., & Sundar, S. (2016). Measuring message credibility: Construction and validation of an exclusive scale. *Journalism & Mass Communication Quarterly, 93*(1), 59–79. https://doi.org/10.1177/1077699015606057

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology, 39*(5), 752–766. https://doi.org/10.1037/0022-3514.39.5.752

Cho, E., Molina, M. D., & 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, 22*(8), 515–520. https://doi.org/10.1089/cyber.2018.0571

Dambanemuya, H. K., & Diakopoulos, N. (2021). Auditing the information quality of news-related queries on the Alexa voice assistant. *Proceedings of the ACM on Human Computer Interaction, 5*(CSCW1), Article 83. https://doi.org/10.1145/3449157

Flanagin, A. J., & Metzger, M. J. (2000). Perceptions of Internet information credibility. *Journalism & Mass Communication Quarterly, 77*(3), 515–540. https://doi.org/10.1177/107769900007700304

Flanagin, A. J., & Metzger, M. J. (2007). The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based information. *New Media & Society, 9*(2), 319–342. https://doi.org/10.1177/1461444807075015

Fogg, B. J., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., Paul, J., Rangnekar, A., Shon, J., Swani, P., & Treinen, M. (2001). What makes Web sites credible? A report on a large quantitative study. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 61–68. https://doi.org/10.1145/365024.365037

Gaiser, F., & Utz, S. (2023). *Data and materials for "The Importance of Modality, Sponsor Credibility and Message Accuracy for perceived message credibility"*. https://doi.org/10.17605/OSF.IO/BQTZM

Gruber, J., Hargittai, E., Karaoglu, G., & Brombach, L. (2021). Algorithm awareness as an important internet skill: The case of voice assistants. *International Journal of Communication, 15*, 1770–1788. https://doi.org/10.5167/uzh-204503

Hu, Y., & Sundar, S. (2010). Effects of online health sources on credibility and behavioral intentions. *Communication Research, 37*(1), 105–132. https://doi.org/10.1177/0093650209351512

Ibelema, M., & Powell, L. (2001). Cable television news viewed as most credible. *Newspaper Research Journal, 22*(1), 41–51. https://doi.org/10.1177/073953290102200104

Ischen, C., Araujo, T., Voorveld, H., Van Noort, G., & Smit, E. (2021, May 27–31). *Is voice really persuasive? The influence of modality in virtual assistant interactions and two alternative explanations*. 71st Annual Conference of the International Communication Association (ICA), Gold Coast, Australia. [Talk]

Jung, E. H., Walsh-Childers, K., & Kim, H.-S. (2016). Factors influencing the perceived credibility of diet-nutrition information web sites. *Computers in Human Behavior, 58*, 37–47. https://doi.org/10.1016/j.chb.2015.11.044

Kalyuga, S., Chandler, P., & Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology, 92*(1), 126–136. https://doi.org/10.1037/0022-0663.92.1.126

Kammerer, Y., & Gerjets, P. (2012). How search engine users evaluate and select web search results: The impact of the search engine interface on credibility assessments. In D. Lewandowski (Ed.), *Web search engine research* (Vol. 4, pp. 251–279). Emerald Group Publishing. https://doi.org/10.1108/S1876-0562(2012)002012a012

Kiousis, S. (2001). Public trust or mistrust? Perceptions of media credibility in the information age. *Mass Communication and Society, 4*(4), 381–403. https://doi.org/10.1207/S15327825MCS0404_4

Lewandowski, D. (2012). Credibility in web search engines. In *Online credibility and digital ethos: Evaluating computer-mediated communication*. https://doi.org/10.4018/978-1-4666-2663-8.ch008

Li, M., & Suh, A. (2021). Machinelike or humanlike? A literature review of anthropomorphism in AI-enabled technology. *Hawaii International Conference on System Sciences 2021 (HICSS-54)*. https://aisel.aisnet.org/hicss-54/in/ai_based_assistants/5

Lüdecke, D. (2021). *Package "sjPlot"*. CRAN. https://cran.r-project.org/web/packages/sjPlot/sjPlot.pdf

Mayer, R. E. (2001). *Multimedia learning*. Cambridge University Press. https://doi.org/10.1017/CBO9781139164603

Metzger, M. J., Flanagin, A. J., Eyal, K., Lemus, D. R., & Mccann, R. M. (2003). Credibility for the 21st century: Integrating perspectives on source, message, and media credibility in the contemporary media environment. *Annals of the International Communication Association, 27*(1), 293–335. https://doi.org/10.1080/23808985.2003.11679029

Metzger, M. J., Flanagin, A. J., & Medders, R. B. (2010). Social and heuristic approaches to credibility evaluation online. *Journal of Communication, 60*(3), 413–439. https://doi.org/10.1111/j.1460-2466.2010.01488.x

Moussawi, S., Koufaris, M., & Benbunan-Fich, R. (2021). How perceptions of intelligence and anthropomorphism affect adoption of personal intelligent agents. *Electronic Markets, 31*(2), 343–364. https://doi.org/10.1007/s12525-020-00411-w

Petty, R. E., & Cacioppo, J. T. (1990). Involvement and persuasion: Tradition versus integration. *Psychological Bulletin, 107*(3), 367–374. https://doi.org/10.1037/0033-2909.107.3.367

Petty, R. E., Cacioppo, J. T., & Goldman, R. (1981). Personal involvement as a determinant of argument-based persuasion. *Journal of Personality and Social Psychology, 41*(5), 847–855. https://doi.org/10.1037/0022-3514.41.5.847

Porcheron, M., Fischer, J. E., Reeves, S., & Sharples, S. (2018). Voice interfaces in everyday life. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 640*, 1–12. https://doi.org/10.1145/3173574.3174214

Powell, T. E., Boomgaarden, H. G., De Swert, K., & de Vreese, C. H. (2018). Video killed the news article? Comparing multimodal framing effects in news videos and articles. *Journal of Broadcasting & Electronic Media, 62*(4), 578–596. https://doi.org/10.1080/08838151.2018.1483935

Qiu, L., & Benbasat, I. (2009). Evaluating anthropomorphic product recommendation agents: A social relationship perspective to designing information systems. *Journal of Management Information Systems, 25*(4), 145–182. https://doi.org/10.2753/MIS0742-1222250405

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. The Center for the Study of Language and Information Publications.

Reinhard, M.-A., & Sporer, S. L. (2010). Content versus source cue information as a basis for credibility judgments. *Social Psychology, 41*(2), 93–104. https://doi.org/10.1027/1864-9335/a000014

Rzepka, C., Berger, B., & Hess, T. (2022). Voice assistant vs. chatbot – Examining the fit between conversational agents' interaction modalities and information search tasks. *Information Systems Frontiers, 24*(3), 839–856. https://doi.org/10.1007/s10796-021-10226-5

Rieh, S. Y., & Belkin, N. (1998). Understanding judgment of information quality and cognitive authority in the WWW. *Journal of the American Society for Information Science and Technology, 35*.

Sundar, S. (2000). Multimedia effects on processing and perception of online news: A study of picture, audio, and video downloads. *Journalism & Mass Communication Quarterly, 77*(3), 480–499. https://doi.org/10.1177/107769900007700302

Sundar, S. (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility. In M. J. Metzger & A. J. Flanagin (Eds.), *Digital media, youth, and credibility* (pp. 73–100). https://www.semanticscholar.org/paper/The-MAIN-Model-%3A-A-Heuristic-Approach-to-Technology-Sundar/de80aa094f380342a632eadb0ee8d4221e8920ba

Sundar, S., Molina, M. D., & Cho, E. (2021). Seeing is believing: Is video modality more powerful in spreading fake news via online messaging apps? *Journal of Computer-Mediated Communication, 26*(6), 301–319. https://doi.org/10.1093/jcmc/zmab010

Sweller, J. (1999). *Instructional design in technical areas*. Australian Council Educational Research.

Tabbers, H. K., Martens, R. L., & Merriënboer, J. J. G. (2004). Multimedia instructions and cognitive load theory: Effects of modality and cueing. *British Journal of Educational Psychology, 74*(1), 71–81. https://doi.org/10.1348/000709904322848824

Tate, M. (2018). *Web wisdom: How to evaluate and create information quality on the web*. https://doi.org/10.1201/b22397

The Smart Audio Report. (2020). *National public media*. https://www.nationalpublicmedia.com/insights/reports/smart-audio-report/

Thompson, S. A. (2022, February 23). Fed up with Google, conspiracy theorists turn to DuckDuckGo. *The New York Times*.

https://www.nytimes.com/2022/02/23/technology/duckduckgo-conspiracy-theories.html

Unkel, J., & Haas, A. (2017). The effects of credibility cues on the selection of search engine results. *Journal of the Association for Information Science and Technology, 68*(8), 1850–1862. https://doi.org/10.1002/asi.23820

Wagner, K., & Schramm-Klein, H. (2019). Alexa, are you human? Investigating anthropomorphism of digital voice assistants – A qualitative approach. *ICIS 2019 Proceedings*, *Vol. 7*. https://aisel.aisnet.org/icis2019/human_computer_interact/human_computer_interact/7

Westerwick, A. (2013). Effects of sponsorship, web site design, and Google Ranking on the credibility of online information. *Journal of Computer-Mediated Communication, 18*(2), 194–211. https://doi.org/10.1111/jcc4.12006

Zuberer, S. (2017). *Digitale Assistenten Bevölkerungsumfrage 2017* [Digital assistants demographic survey 2017]. PwC Communications. https://docplayer.org/53107384-Digitale-assistenten-bevoelkerungsbefragung.html

History

Conflict of Interest

We have no known conflict of interest to disclose.

Open Data

The authors have pre-registered this research with an analysis plan, which is retrievable at https://aspredicted.org/2y8er.pdf and https://aspredicted.org/vb4u2.pdf.
All data and materials are available at https://osf.io/bqtzm/ (Gaiser & Utz, 2023).

Funding

ORCID
Franziska Gaiser
https://orcid.org/0000-0001-9841-4102
Sonja Utz
https://orcid.org/0000-0002-7979-3554

Franziska Gaiser
Leibniz-Institut für Wissensmedien
Schleichstraße 6
72076 Tübingen
Germany
f.gaiser@iwm-tuebingen.de

Franziska Gaiser is a PhD student at Leibniz-Institut für Wissensmedien in Tübingen. Her main research focuses on human-machine communication, especially smart speakers and how users perceive and evaluate information from them.

Sonja Utz (PhD) is the head of the everyday media lab at Leibniz-Institut für Wissensmedien in Tübingen and a full professor for communication via social media at University of Tübingen. Her research focuses on the effects of social and mobile media use, especially in knowledge related contexts, as well as on human-machine interaction.