Examining the Use of Voice Assistants: A Value-Focused Thinking Approach

Completed Research
Christine Rzepka
LMU Munich
rzepka@bwl.lmu.de

Abstract
Speech interaction is becoming increasingly sophisticated and prevalent in everyday life due to recent advances in the area of artificial intelligence and high adoption rates of voice assistants (VAs). However, besides using a VA, individuals typically have various alternatives when pursuing certain tasks. To better understand when and why users choose speech interaction over traditional user interfaces, we conducted 31 one-to-one interviews. We analyzed the benefits and costs that users evaluate when using VA by applying means-end chain theory and value-focused thinking. Our findings reveal five fundamental (efficiency, convenience, ease of use, less cognitive effort, and enjoyment) and twelve means objectives that users evaluate when deciding on whether to use their VA or not. The resulting means-ends objectives network informs researchers by providing a comprehensive overview of relevant factors regarding VA use. Practitioners may learn how to further improve and sell VA characteristics.

Keywords
Speech interaction, voice assistant, means-end chain theory, value-focused thinking.

Introduction
Natural language processing and speech recognition have become more and more sophisticated due to increased computing power, large availability of linguistic data, improved machine learning methods, and a better understanding of human language (Hirschberg and Manning 2015). By making use of these technological advancements, voice assistants (VAs) such as Amazon’s Alexa or Apple’s Siri can process spoken input and provide output as synthetic voice (Hoy 2018). This allows users to interact with VAs in a natural and conversational manner. VAs can be integrated in existing devices such as smartphones, which are usually operated via touchscreens, or are embedded in smart speakers, which can be operated by voice commands only. Overall, VAs offer users a new way (i.e. speech interaction) to carry out tasks in using information systems. The increasing diffusion of smart speakers (Voicebot.ai 2018) illustrates the adoption of speech interaction based on VAs in private as well as organizational contexts. However, while VAs enable users to carry out tasks in a different way, these tasks could usually also be solved using traditional user interfaces (e.g. setting a timer, requesting the weather forecast, making a purchase). This raises the question what drives users to choose speech interaction over other modes of interaction with information systems.

Until now, the main benefits of speech interaction are seen in time savings and hands-free and eyes-free use (Cohen and Oviatt 1995; Moussawi 2018). However, ambiguous results on, for example, efficiency gains (Le Bigot et al. 2004) suggest that they are highly dependent on users’ tasks, contexts, and own characteristics. Extant research implies that speech interaction is particularly useful for tasks of low complexity (Zamora 2017), that require multitasking (Luger and Sellen 2016), or in situations where other input modalities are less usable (Cohen and Oviatt 1995). Beside those advantages, speech interaction is also prone to various adoption barriers such as privacy concerns and social acceptability (Easwara Moorothy and Vu 2015), or speech recognition errors (D’Mello et al. 2010). From a user’s perspective, speech interaction requires higher mental workload compared to text (Le Bigot et al. 2004), which may be explained by the unnaturalness at its early stages (e.g. through isolated commands) (Bernsen and Dybkjaer 1999). With more and more sophisticated conversational capabilities, however, VAs may provide new
values to the individual. Therefore, we seek to provide an integrated and comprehensive perspective on the values of using VAs. Following the call of Liu et al. (2017) for more behavioral research on speech interaction in the IS domain, we hence pose the research question:

RQ: What are the values that users seek to achieve by using voice assistants?

To answer this research question, we draw on means-end chain theory to describe relationships between VAs’ attributes, the resulting consequences for the users, and the personal values that users derive from those consequences (Gutman 1982, 1991). We apply value-focused thinking (VFT) to identify and organize the users’ personal values (Keeney 1992). VFT allows to analyze how these values can be achieved and to systematically represent the relationships between the values and their determinants in a means-ends objectives network. Following Keeney’s (1992) guidelines for conducting VFT, we have interviewed 31 VA users in one-to-one interviews and identified five fundamental and twelve means objectives.

This paper is structured as follows: First, the theoretical background covers extant research on speech interaction, VAs and means-end chain theory. Second, VFT as research method and the data collection and analysis procedure are described. Third, we present and discuss findings from the interviews, compare them to extant research, and provide implications for both researchers and practitioners.

Theoretical Background

Speech Interaction and Voice Assistants

With speech being the primary and most natural way of communication for humans, efforts to build systems that can understand spoken language and respond via voice have a long history. The first successful systems to automatically recognize speech date back to the early 1950s, e.g. by Davis et al. (1952). Milestones in the development of automatic speech recognition include the use of pattern recognition methods in the 1960s and later the application of statistical methods such as hidden Markov models and neural networks that enabled those systems to understand medium to large vocabularies of more than 1000 words (Juang and Rabiner 2005). In the 1990s, initial systems were able to interact in conversations with their users through full semantic models and the addition of text-to-speech synthesis. Those so-called spoken dialogue systems minimally consist of an automatic speech recognizer that interprets human speech, a dialogue manager that determines and performs the requested action, and text-to-speech synthesis that responds in a spoken form (Glass 1999). With the launch of Apple’s Siri in 2011, spoken dialogue systems have evolved from simple question-answering systems toward more sophisticated VAs that obtain knowledge from users’ speech and usage data in order to adapt to their behavior and improve service quality by leveraging collective intelligence (Knote et al. 2019). Speech interaction based on VAs may thereby either be voice-based, i.e. having a single modality (e.g. Amazon Echo), or voice-enhanced, i.e. having a multimodal interface (e.g. Google Assistant). While the former is predominantly included in standalone smart speaker applications with speech as the only interaction mode, the latter is added to existing devices such as smartphones, laptops, or televisions that provide supplementary visual output and other interaction modes. Current VAs assist the user to carry out tasks via speech, primarily in simple information retrieval or service execution tasks in the private context (Gnewuch et al. 2017; Knote et al. 2019).

However, information systems (IS) research has so far predominantly focused on designing and improving prototypes for speech recognition, while individual-level research on specific applications such as VAs is still scarce (Liu et al. 2017). Research from related disciplines including linguistics, psychology, and computer science suggest that speech differs from other interaction modes such as text in terms of acquisition, production, transmission, and reception (Akkinaso 1982). Using speech to interact with machines provides benefits in situations where hands and eyes are occupied and when there is only a limited keyboard (Cohen and Oviatt 1995), whereas text supersedes for complex activities that require a confirmation and hence control on the user’s side (Zamora 2017). Furthermore, using voice instead of text promotes a warmer attitude and more personal dialogues (Novielli et al. 2010) and can also influence users’ likelihood to trust machines with their personal information (Schroeder and Schroeder 2018). While extant research often reports on higher efficiency and faster task performance of speech than text (Cohen and Oviatt 1995; D’Mello et al. 2010), other authors contradict these results with higher efficiency and less mental workload for text compared to speech in an information search task (Le Bigot et al. 2004). These findings illustrate benefits and shortcomings on speech interaction in general. Initial research specifically
in the VA context further confirms the value of trust (Nasirian et al. 2017) and privacy (Easwara Moorthy and Vu 2015) as determinants of VA adoption, and reveals users’ expectations that are so far not aligned with the capabilities of state-of-the-art VAs (Luger and Sellen 2016; Moussawi 2018). However, there is until now no comprehensive understanding of what utility is derived from VAs and their increasingly natural and humanlike conversational characteristics.

**Values and Means-End Chain Theory**

A value is defined as a “belief that a specific mode of conduct or end-state of existence is personally or socially preferable to an opposite or converse mode of conduct or end-state of existence” (Rokeach 1973, p.5). Hence, it describes a preference for modes of behavior (e.g. honest) or end-states of existence (e.g. happiness) (Gutman 1982). In the decision-making context, values constitute principles that users try to achieve and what they consider when evaluating a decision, e.g. ethics, desired characteristics, or attitudes (Keeney 1992). Means-End Chain (MEC) theory understands values as being at the higher levels of abstraction within individuals’ hierarchical cognitive structures, while an object’s attributes are the lowest levels (Gutman 1982). Based on this assumption, the theory describes the relationships between an object’s attributes (lower-level “means”), its direct consequences (higher-level “means”), and the personal values (“ends”) that users derive from those consequences (Gutman 1991). Means are defined as products or activities that people use or do, while ends describe the valued states of being. Consequences at the lower levels facilitate or cause the achievement of valued states. These consequences are physiological, psychological, or sociological results from consumers’ behavior and may be desirable or undesirable. Desirable actions are preferred, whereas undesired consequences should be minimized (Gutman 1982).

Originating from marketing research, MEC theory has been applied in various academic disciplines. Studies in the IS domain have built hierarchical value maps to depict the relevance of communication features (Köster 2016), integrated the theory with the Technology Acceptance Model (TAM) to explain which factors affect document management system use (Chiu 2005), and applied VFT to understand values of mobile technology in education (Sheng et al. 2010). Since VAs have certain capabilities that other technologies do not have, i.e. natural language processing, communicating, and learning, they are able to mimic humanlike behavior to a certain extent (Russel and Norvig 2010). By applying MEC theory, we aim to understand users’ individual decision-making processes based on the linkages between the VA’s characteristics and its consequences that directly affect users’ values. These values may influence users’ intention to use the system, analogous to factors known from TAM (Davis et al. 1989) and UTAUT2 (Venkatesh 2012). However, VA’s technology characteristics may also change users’ perceptions known from those models (Nguyen and Sidorova 2017). Hence, MEC theory provides a conceptual base to examine VA use behavior in depth by also taking into account VAs’ characteristics. Through the application of VFT as methodological approach, we are further able to not only depict hierarchical representations of users’ value chains, but also to present desirable and undesirable consequences. VFT may be used for any decision that is worth thinking about, whether in organizational or private contexts, as it provides a structured approach to identify what a decision maker cares about in a specific situation (Keeney 1992). VFT is therefore a suitable method to understand why users’ decide to use VAs and provides implications for the design of future VAs.

**Methodology**

Drawing from the conceptual foundations of MEC theory, we apply VFT to identify and organize user values (Keeney 1992). The main objective of VFT is to identify what fundamental values are important to the user and how these values can be achieved. Thereby, values are made explicit by objectives, which are divided into means and fundamental objectives: means objectives describe the attributes from the lower and higher levels known from MEC theory and have implications for other objectives, while fundamental objectives define users’ personal values and essential reasons for interest in the decision situation. Hence, means objectives directly influence the fundamental objectives, but may also influence each other. Their interrelationships are depicted in a means-ends objectives network. VFT has already been successfully applied in the IS domain, e.g. in understanding users’ values of mobile technology (Sheng et al. 2010), internet commerce (Torkzadeh and Dhillon 2002), and social networks (Dhillon and Chowdhuri 2013). Since VFT has proven useful to uncover fundamental values that are often hidden and provides a systematic way to identify relationships between the underlying objectives, this method is appropriate to identify what users seek to achieve by using VAs. After data collection, we apply VFT following the steps proposed by.
Keeney (1992): first, an initial list of objectives is derived from the data and converted into a common form; second, the objectives are structured into fundamental and means objectives; third, a means-ends objectives network that depicts the relationships is built. Following, the process is described in detail.

**Data Collection**

As our research question focuses on the objectives that users seek to achieve by using VAs, only individuals who have already used VAs and would consider themselves as ‘regular’ users were selected for interviews to ensure that they had a good understanding of the capabilities, benefits, and limitations. We applied convenience sampling and recruited participants from all ages via announcements in lectures for senior students, university mailing lists, and public places. Following Keeney’s (1992) guidelines on VFT, we conducted one-to-one interviews between November 2018 and February 2019 with 31 VA users in person or via telephone. The interviews lasted 25 minutes on average. Theoretical saturation was reached after the 16th interview, the following 15 interviews confirmed that we captured all individual objectives. The resulting sample includes 14 female and 17 male participants, aged between 20 and 77 years with a mean age of 29. Participants most frequently used Amazon’s Alexa (74%), followed by Apple’s Siri (23%), and the Google Assistant (23%) for playing music, asking questions, and smart home control. Detailed information on the participants are depicted in Table 1. Please note that participants may have used more than one VA.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Specification</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>20-39</td>
<td>25</td>
<td>81%</td>
</tr>
<tr>
<td></td>
<td>40-59</td>
<td>5</td>
<td>16%</td>
</tr>
<tr>
<td></td>
<td>&gt; 59</td>
<td>1</td>
<td>3%</td>
</tr>
<tr>
<td>Gender</td>
<td>Male</td>
<td>17</td>
<td>55%</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>14</td>
<td>45%</td>
</tr>
<tr>
<td>Profession</td>
<td>Student</td>
<td>13</td>
<td>42%</td>
</tr>
<tr>
<td></td>
<td>Employee</td>
<td>16</td>
<td>52%</td>
</tr>
<tr>
<td></td>
<td>Others</td>
<td>2</td>
<td>6%</td>
</tr>
<tr>
<td>VA type</td>
<td>Amazon’s Alexa</td>
<td>23</td>
<td>74%</td>
</tr>
<tr>
<td></td>
<td>Apple’s Siri</td>
<td>7</td>
<td>23%</td>
</tr>
<tr>
<td></td>
<td>Google Assistant</td>
<td>7</td>
<td>23%</td>
</tr>
<tr>
<td></td>
<td>Others</td>
<td>3</td>
<td>10%</td>
</tr>
<tr>
<td>(Average) VA use</td>
<td>Daily</td>
<td>23</td>
<td>74%</td>
</tr>
<tr>
<td></td>
<td>Weekly</td>
<td>6</td>
<td>19%</td>
</tr>
<tr>
<td></td>
<td>Monthly</td>
<td>2</td>
<td>6%</td>
</tr>
</tbody>
</table>

**Table 1. Information on Participants**

To elicit users’ mean and fundamental objectives, two different interview techniques were used: “benefits and problems” and “wish list” (Keeney 1992). Since our participants were regular users of VAs, “benefits and problems” encouraged discussions about their experiences. “Wish list” was chosen as promising technique to account for future technological advancements and functional limitations of state-of-the-art VAs. Building on those techniques, probing questions in the beginning of the interviews were: “What are the benefits of using VAs?” “What problems or concerns can arise in using VAs?” “If there is no limitation at all, what are the features or functions of VAs you wish to have?”. These questions elicit the initial list of values of using VAs. To get a deeper understanding of those, the laddering interview technique has proven useful (Sheng et al. 2010). Laddering describes an in-depth interviewing technique, which helps to understand the process in which users translate attributes into values (Reynold and Gutman 1988). Probing questions such as “Why is that important to you?” determine the relationships between attributes, consequences, and values. To encourage further thinking on additional aspects, we then asked participants about their opinions on different system characteristics such as the system’s input and output modalities, typical tasks that VAs are used for, and contexts in which the user may decide between using a VA or another interaction mode. This approach aims to assess the decision context from different perspectives and to foster further critical evaluation of users’ hidden objectives (Keeney 1999). The probing questions were repeated until no further values were identified. In the end, users’ demographics were obtained.
Data Analysis

Two researchers coded six of the literal transcripts of the interviews in-vivo to derive a list of means and fundamental objectives based on users’ expressions (Miles et al. 2014). After discussing mismatches and reaching consensus on the naming of concepts, one researcher coded the rest of the interviews. Redundancies were then eliminated, but similar objectives remained and were clustered to build thematic categories. Each objective was converted into a common form, consisting of a decision context, an object, and a direction of preference. The decision context was framed as whether or not using VAs. In this decision context, the object is a noun and the direction of preference is described by a verb. In this way, statements such as “it is quicker to talk” were converted into “maximize efficiency”. Similar objectives were further clustered into overarching categories. Finally, means-ends chain relationships were derived from the interviews by asking the “Why is That Important” test. For example, if one participant stated: “I don’t want to push a button to talk to Siri”, the researcher asked “why”. The answer was then: “because I want to have my hands free when talking to Siri”, the question was again: “why is that important”, and the answer was: “because then I can do other things simultaneously and save time”. As a result, we provide a means-ends objectives network that represents users’ values of using VAs and their relationships in a systematic way.

Findings

After data analysis, we derived a list of fundamental and means objectives (see Table 2). Efficiency, convenience, ease of use, minimal cognitive effort, and enjoyment are the fundamental objectives (i.e. users’ personal values and essential reasons) that maximize users’ overall value of using VAs. They are directly affected by twelve means objectives that provide information on how the fundamental values may be achieved. The bullet points in the table show examples of these objectives. Please note that the presentation in Table 2 does not follow a specific order, as the intent of this work is to provide a comprehensive list of objectives without eliciting a hierarchy. However, to increase transparency, the frequencies of participants citing the fundamental objectives are provided in the text. Fundamental objectives are highlighted in bold, while associated means objectives are italic in the descriptions.

<table>
<thead>
<tr>
<th>Fundamental Objectives</th>
<th>Means Objectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximize efficiency</td>
<td>Maximize convenience</td>
</tr>
<tr>
<td>- Ensure faster task completion</td>
<td>- Minimize physical effort</td>
</tr>
<tr>
<td>Maximize ease of use</td>
<td>Minimize cognitive effort</td>
</tr>
<tr>
<td>- Ensure easy access to underlying system</td>
<td>- Minimize deliberate thinking</td>
</tr>
<tr>
<td>Maximize enjoyment</td>
<td></td>
</tr>
<tr>
<td>- Maximize joy of the interaction itself</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Ensure hands-free and eyes-free use</td>
<td>Minimize speech recognition errors</td>
</tr>
<tr>
<td>- Maximize multitasking</td>
<td>- Minimize repetitions of commands</td>
</tr>
<tr>
<td>Maximize naturalness of conversation</td>
<td>Maximize system transparency</td>
</tr>
<tr>
<td>- Minimize ‘command style’ interaction</td>
<td>- Maximize user understanding of the system</td>
</tr>
<tr>
<td>Ensure offline functionality</td>
<td></td>
</tr>
<tr>
<td>- Minimize dependence on internet connection</td>
<td>Maximize compatibility</td>
</tr>
<tr>
<td></td>
<td>- Maximize integration of different devices &amp; apps</td>
</tr>
<tr>
<td>Maximize system adaptation</td>
<td></td>
</tr>
<tr>
<td>- Maximize learning from users’ habits</td>
<td>Maximize trust</td>
</tr>
<tr>
<td></td>
<td>- Ensure that the system does what is expected</td>
</tr>
<tr>
<td>Ensure privacy</td>
<td></td>
</tr>
<tr>
<td>- Minimize insecurity about data storage &amp; usage</td>
<td>Maximize social acceptability</td>
</tr>
<tr>
<td></td>
<td>- Minimize awkwardness of speaking to computer</td>
</tr>
<tr>
<td>Provide visual output for complex tasks</td>
<td>Provide spoken confirmation for complex tasks</td>
</tr>
<tr>
<td>- Provide list of choice alternatives</td>
<td>- Provide summary of what the system did</td>
</tr>
</tbody>
</table>

Table 2. Fundamental and Means Objectives of Using VAs

Efficiency was stated by 25 interview participants as being a fundamental objective of using VAs. These users would only decide to speak to their VA if they assume that speaking will be faster than any other alternative to fulfill the task. This objective is mainly achieved by the opportunity to interact with the system without using hands or eyes, therefore allowing the user to do other things simultaneously: “I feel like I
lose the benefit when I have to interact manually, even if it is just pressing a button” [P19]. In addition, speech recognition errors counteract users’ efficiency goals: “I always use Alexa while doing something else and when she does not understand me, I repeat it, and then she doesn’t understand me again. Then I stop doing what I did and approach her directly to repeat it slowly and clearly. In the end, it would have just been faster to do it myself” [P5]. Users also state that errors are especially severe when there is no internet connection, as there is no offline support. Speech recognition errors further affect users’ trust in the system to be able to conduct certain tasks for them, especially if there is some kind of risk associated, e.g. financial or work related. This is why users would like to get a spoken confirmation on tasks for which the successful execution is not directly observable (e.g. as is the case for light control): “For me it is important to get a confirmation. I mistrust the system to understand me and therefore would like to be sure. Especially for getting up, as the consequences would be severe when I miss a meeting” [P2]. At the same time, users fear that the assistant might not act in their sense, e.g. order a product that they did not want.

Another fundamental objective of using VAs is convenience, which was mentioned by 20 participants. As users do not have restrictions regarding their distance to the system and can use it without hands and eyes, users are able to minimize their physical effort. To further maximize their convenience goals, users wish that their VA may be integrated with diverse features and applications, as limited compatibility forces users to perform their routine tasks on their own: “I would like to do everything in my house with Alexa” [P27] or “I also want to control other applications on my smartphone via my VA” [P3]. Participants also want the VA to learn from them and adapt its behavior to their habits: “I want her to know more than my laundry automat now. I want her to know my usual settings and that she uses them automatically without me getting up or while I am doing something else. It would be more convenient” [P20].

Compared to interacting with a system by other means, 17 interviewees value their VAs for their ease of use. Various participants find that using their VAs is easier than using a touchpad, keyboard, joystick, or buttons: “Because it is easier than typing every word on this small keyboard” [P14]. They mainly criticize low usability of those interfaces because of complex input mechanisms or the particular design of the interface (e.g. buttons that are too small to be used easily). Similarly, other participants preferred VAs because of confusing application menus and user experience designs. To maximize users’ perceived ease of use, participants wished for a more natural and intuitive way of conversation as they would often not know the right commands for their desired task: “It annoys me, I always do it wrong. Usually, I use the term ‘turn off’ to signal that I don’t want to hear music anymore, not ‘stop the music’. It is different in everyday language” [P20]. Related to this unnaturalness of conversation, users would also not use their VA in public or crowded places as they feel it would not be socially accepted. Users feel awkward when talking to their smartphone in front of strangers and also do not want to disturb them.

Related to users’ perceived ease of use, 13 users in our sample also value that they experience less cognitive effort through speaking to their VA as it is more intuitive and less deliberate than any other form of interaction. This benefit is particularly strong when comparing VA use to text input for reminders or messages as they do not have to explicitly think about grammar or syntax: “You don’t have to take your phone in your hand and think about what you’d like to write in your calendar. You can talk and it reminds you of what you said” [P3]. Also, users say that using voice is a welcome variation to using their smartphone or computer as it minimizes distraction from other information presented on the screen when pursuing a certain task: “It reduces my screen time. I don’t have to look at the screen all the time. […] And the other thing is that the smartphone is also a distraction. I might see a mail and can’t stop but read it” [P18]. Again, to achieve this goal, users require the VA to be used hands- and eyes-free, as this benefit cannot be achieved otherwise. This is however only feasible for routine and simple tasks that only require a short and unambiguous answer from the system. With rising task complexity and choice alternatives, users wish for visual representations to shorten the length of the VA’s reply and reduce the need to memorize every choice.

Furthermore, eight participants value that speaking to their VA maximizes their enjoyment. They are curious about the assistant’s reply and enjoy talking to it because of two reasons: first, they simply like exploring a new technology and its features and behavior: “It is fun to talk to the system and to realize that it understands you. I think it’s cool, I like innovative things” [P1]; second, they specifically have fun using the entertainment functions: “It’s a game, for personal entertainment” [P18]. Note that while the former is true for all VA functions as this value is derived from the users’ technical affinity of using speech technology itself, the latter refers to specific situations in which users take the time to interact with their VA for hedonic instead of utilitarian reasons. Consequently, they do not care about any efficiency gains in these situations.
However, they also do not see enjoyment as their primary reason for using VAs in general: “If Alexa would not have all those useful features, I wouldn’t use her. But if she would not have any entertainment features, I would still use her” [P20]. Although not only related to users’ enjoyment, also privacy concerns affect users’ value perceptions of VAs. They feel insecure about not knowing when their VA listens to their conversations. Therefore, participants state that they placed their system deliberately at a place where they would not talk about sensitive topics, such as the bathroom or kitchen. They further stress the necessity to keep some sense of control through maximizing system transparency, i.e. having a visual indication for audio recordings: “I don’t know exactly how it works, but the microphone listens the whole time and analyzes whether I said ‘Echo’. And when it recognizes ‘Echo’, the light flashes and the audio recording turns on, which will then be sent to Amazon. This means it is extremely important for me to see when this happens” [P18]. Users further expect the provider to communicate the data location and utilization more transparently: “I have the feeling that I don’t have any control over my data. I find this very negative” [P4].

The relationships of fundamental and means objectives are depicted in a means-ends objectives network in Figure 1 (note that task-dependent values are shown in grey boxes). The fundamental objectives are shown on the right, while the means objectives and their interrelationships are depicted on the left.

![Figure 1. Means-Ends Objectives Network for Using VAs](image)

**Discussion**

In line with previous research, our findings confirm the relevance of existing theoretical models within the IS discipline in the VA context. Specifically, users’ fundamental objectives of maximizing ease of use, minimizing cognitive effort, and maximizing enjoyment for using VAs relate to existing constructs of “effort expectancy” and “hedonic motivation” in UTAUT2 (Venkatesh 2012) and “perceived ease of use” in TAM (Davis et al. 1989). Regarding users’ “performance expectancy” or “perceived usefulness”, our findings further specify the actual benefits that users expect from VA use, i.e. efficiency and convenience. While those constructs seem particularly relevant for the use of this technology, others are less important. For example, our participants gave counterintuitive answers to the construct of “social influence”: although they said that their friends would never use a smart speaker because of privacy concerns and risk perceptions, it did not affect their use behavior. Some participants explained this with their technical experience, as they claimed to have a better understanding of the technology and are therefore able to assess the risks.
Our results are also consistent with research on speech interaction and VAs. Users’ efficiency goal confirms prior literature on users’ expected time savings from using voice (Moussawi 2018) and empirical findings that show initial evidence for achieving this goal in the tutoring context (D’Mello et al. 2010). However, previous results also imply that voice is less efficient than text in an information search task (Le Bigot et al. 2004), which might be due to different task characteristics. Although most of our participants agreed on their efficiency goal, some of them also explained differences in expected time savings dependent on the tasks they pursue. For example, if users do not specifically know what they search for and expect turn-taking with the system to specify requirements, they expect longer interactions and prefer visual displays to facilitate their choice. Our research further supports the negative impact of speech recognition errors that has already been observed for tutoring (D’Mello et al. 2010) and for voice search (Jiang et al. 2013).

Furthermore, our results stress the relevance of trust, privacy, and social acceptability as potential inhibitors for maximizing the values of VAs. While trust is not a new phenomenon in IS research and has already been studied in the context of VAs (Nasirian et al. 2017), previous research did not account for differences between technology and interpersonal trust. Similar to previous findings in the social media context, our participants also differentiate between trust in the underlying technology (e.g., its functionality) and trust in the provider (e.g., benevolence) (Lankton and McKnight 2011). Regarding users’ privacy concerns, constructs from previous literature are also supported such as improper access, unauthorized secondary use, collection (Smith et al. 1996), awareness, and control (Malhotra et al. 2004). Although initial studies investigated privacy concerns for VAs (Easwara-Moorthy and Vu 2015), further research is needed, especially in the context of individuals’ non-adoption behavior.

**Conclusion**

This study set out to identify, define, and understand what users value and try to achieve by using their VA. Typically, individuals have various decision alternatives when pursuing a certain task: they may either interact through a traditional interface, or they may use a VA. To understand users’ decision-making processes, we conducted one-to-one interviews and analyzed the data using VFT. This approach allowed us to identify users’ fundamental objectives as well as the means objectives that directly affect the achievement of the former. As a result, the means-ends objectives network summarizes the identified values and their relationships. We find that users derive both utilitarian and hedonic benefits from VAs, whereby the utilitarian values outweigh the hedonic ones. Furthermore, our results confirm the importance of users’ trust and privacy concerns in the context of VAs, which needs further investigation. Therefore, also differences between different tasks and task characteristics need to be considered as our data also shows that users’ values differ according to varying degrees of task complexity and risk.

Although we carefully followed the guidelines of VFT, this study has several limitations that provide avenues for future research projects. First, our convenience sample could have been extended. Although we interviewed VA users from all ages, the ages were not equally distributed with the majority being aged between 20 and 30 years. Extending the sample would further allow for subgroup analysis in regard to user characteristics such as age, gender, and use frequency and experience. Second, we have only interviewed regular VA users. Although this approach ensured that we could gather detailed insights from experienced users, additional insights from non-adopters may provide a better understanding for their non- adoption decision and perception of VAs. Third, our study focuses on user values of VAs in the private context. Future research projects may investigate the values of VAs in the organizational context or specify the context and conduct more detailed analyses on the interrelationships between users’ values.

This study contributes to IS research by providing in-depth analysis of means-ends value chains that explain users’ decision-making processes regarding an increasingly adopted type of human-computer interaction: speech interaction. Thereby, our contribution touches three areas: first, our study is one of the first applications of VFT in the private context within the IS discipline. Since Barclay (2014) stresses the potential of this methodological approach for IS research and Keeney (1992) encourages to extend this thinking also to the individual level, we answer those research calls for VAs and provide an example for further applications. Second, by discussing our findings against extant literature, we identify research gaps and avenues for future projects. Through the depiction of the relationships between users’ means and fundamental objectives, we reveal connections that have not been investigated before and may provide interesting extensions for quantitative models. For example, extant research lacks explanations on how the relation of voice-input and output affects users’ cognitive effort in using VAs. Although developers work
on improving conversational capabilities of VAs, our findings suggest that users may not want the system to answer in a human way. Also, quantitative projects may investigate the impact of system transparency on users’ perceptions, such as their privacy concerns that could be reduced according to our results and users’ trust that may be increased. Third, our study stresses the relevance of task and context characteristics for the applicability and usefulness of speech interaction. Although this relation has already been shown in other contexts and applications of the Task-Technology Fit Model (Goodhue and Thompson 1995), this has so far not been investigated for VAs. Furthermore, this research will inform developers and providers of VAs about the most important attributes and values to the user. This helps providers and marketers of VAs to formulate value propositions to their customers. In addition, our interviews reveal current problems and shortcomings and future wishes regarding VAs, which may help to design better speech-based applications. However, also providers of conventional hardware products may use these insights to think about the potential of adding VAs to their interfaces and whether the use cases are suitable for VAs.

REFERENCES

Examining the Use of Voice Assistants


Twenty-fifth Americas Conference on Information Systems, Cancun, 2019 10