

Is It Human? The Role of Anthropomorphism as a Driver for the Successful Acceptance of Digital Voice Assistants

Katja Wagner
University of Siegen
katja.wagner@uni-siegen.de

Frederic Nimmermann
University of Siegen
frederic.nimmermann@uni-siegen.de

Hanna Schramm-Klein
University of Siegen
hsk@uni-siegen.de

Abstract

The market of digital voice has grown significantly over the recent years. Big players like Amazon, Google, Apple, Microsoft and Samsung are focusing on the development and expansion of their assistants. Especially smart speakers are on the rise but also in smartphone integrated voice applications are getting more popular. The main characteristics of this new technology are both elements of human-computer-interaction and especially the attribution of human characteristics. Although there is an increase of the number of current users as well as of consumers intending to use digital voice assistants in the future, drivers and barriers of digital voice assistants have not yet been sufficiently empirically investigated, especially concerning the phenomenon of anthropomorphism. This study points to additional key factors that are important to foster broader acceptance. Our empirical study is based on the UTAUT2 and highlights the importance of anthropomorphism in relation to other determinants known from the literature.

1. Introduction

Digital voice assistants, also referred to as conversational agents, are revolutionizing our access to web content and our use of technology, e.g., of smart home devices. In the first three quarters of 2017, more than 17 million smart speakers were delivered worldwide and another 16 million during the holiday season [32]. This development means a massive shift in the usage and reception behavior of web content. Experts estimate that by 2020, half of online searches worldwide will be made by voice [34]. Moreover, a study of Capgemini, one of the global leaders in consulting and IT services, revealed that already 51% of the questioned people use this technology [15]. One possible reason for the success is rooted in consumer behavior: As stated by Tadeusiewicz, language is the most natural and comfortable kind to communicate [62].

By definition, conversational agents are systems whose purpose is to provide certain services to the user, in a manner that is modeled on interpersonal

interactions, to provide the highest level of naturalness and convenience to achieve comfortability, wherein the control of the system happens via speech [27]. As they usually are systems of artificial intelligence (AI), conversational agents are also referred to as intelligent personal assistants (IPAs). Thus, an IPA is an application that uses inputs such as the user's voice, images, and contextual information to assist in answering natural language questions, as well as making recommendations and performing actions [27]. According to Hyes-Roth, intelligent assistants are characterized by certain skills such as determining actions, solving problems and drawing inferences [28]. However, following the definition of McCarthy, one of the pioneers in the field of artificial intelligence [1], the goal of AI is to develop machines that behave as if they have intelligence, it is currently controversial whether voice assistants fit in this category. Conversational agents cannot imitate human intelligence, as the systems do not show behavior on their own initiative, but simulate it based on given patterns. Contrary to criticism, however, it is argued that voice assistants will feel more and more human to the users over time, to the point where users can no longer recognize the difference between man and machine because the assistants move in roughly the same interactive paradigm [14].

The extent to which the interpersonal interactions positively influences the intention to use can be investigated by analyzing the relevance of anthropomorphism. Anthropomorphism is defined as the tendency to attribute the actual or perceived behavior of non-human actors, human characteristics, motivations, intentions or emotions [18]. According to the results of Epley et al., anthropomorphizing of non-human actors is based on two fundamental causes: First, as a social being, humans are always in search of interaction with other people. Second, the classification of environmental influences helps to understand them and to keep them in control [58].

In general, the considered elements can be divided into two areas: the relevance of functional components, and the phenomenon of personification of technical devices, the so-called anthropomorphism. These are

assumed to complement relevant determinants in the acceptance of digital voice assistants. However, most literature is concerned with embodied conversational agents (ECA), in which anthropomorphism usually plays a key role in influencing [7, 23, 24]. The use of natural language in these studies, however is usually of minor importance compared to the psyche presence of ECAs. Furthermore, many studies related to conversational agents deal with technological aspects of the software [9, 26, 28]. Some treatises refer to the effect of personification or the integration of emotions in the design of conversational agents, but without empirically related to the user acceptance [13] or they discuss the topic in the form of an overview [19]. Surprisingly, as far as we know, no study has provided empirically insights into anthropomorphism concerning non-ECAs and adequately addressed management implications to enhance the acceptance of digital voice assistants. In sum, we argue that recent research indicates that interaction with robots or assistants, for instance, is not just a pure interaction with technology [37, 55]. For instance, a recent study shows that vocal interaction can actually trigger emotions [30]. Hence, there are consequences of adding human characteristics to machines. The results and the questions raised by these studies seem to us sufficient to deal with the topic of anthropomorphism in more depth.

According to the Uncanny Valley Theory where humanlike robots are only evaluated as positive to a certain degree [44, 45], it is interesting to investigate whether this phenomenon can also be related to digital voice assistants. Furthermore, the computers are social actors paradigm (CASA) implies that computers are assigned similar attributes as humans [47]. Here, we see a gap in literature: While these elements seem to play a role, to our knowledge no study discusses these central points as research questions: (1) *Which role plays anthropomorphism concerning the behavioral intention to use voice assistants?* (2) *Which investigated factor of anthropomorphism influences the behavioral intention the most?* (3) *What are in general further relevant*

factors in the consumers' behavioral intention of using this technology?

To answer these questions, we use the UTAUT2 in terms of digital voice assistants by enlarging the existing model by adding elements of anthropomorphism. Hereby we contribute to a more profound understanding of technology acceptance by respecting anthropomorphism as a novel perceptual dimension for voice assistance in technology. Thus, based on this more holistic approach of technology acceptance for voice assistants we derive implications for further research and managerial implications.

2. Theoretical framework and hypotheses development

Based on literature, we use the UTAUT2 to explain the acceptance of voice assistants. However, we believe based on the CASA paradigm and contrary to the uncanny-valley-paradox that the effect of anthropomorphism, i.e., more humanlike, plays an essential part in the acceptance of voice assistants. Hence, we include anthropomorphism factors in comparison to classical technology acceptance factors in relation to the positive behavioral intention to gain a deeper understanding of important humanlike features. Our conceptual framework is summarized in Figure 1.

2.1. The determinants of the UTAUT2 for digital voice assistants

In the recent decades of technology acceptance research, a variety of theoretical models for explaining technology acceptance and usage has been developed. In their work, Venkatesh et al. established theories of technology acceptance research into a more comprehensive model, resulting in the UTAUT2 [65]. Hence, we adopt these determinants to the acceptance of digital voice assistants: performance expectancy (PE), effort expectancy (EE), hedonic motivation (HM), price value (PV), habit (H), social influence (SI), facilitating conditions (FC), and behavioral intention (BI). Additionally, we include anthropomorphism factors: likeability (L), perceived sociability (PS), animacy (A), and humanlike-fit (HF).

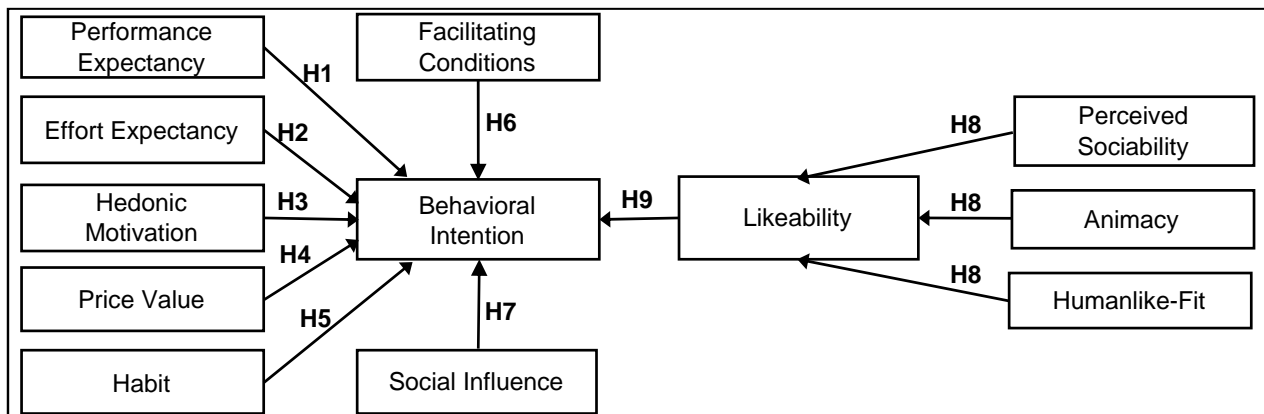


Figure 1. Theoretical Framework

(HEDMOT), price value (PV), habit, facilitating conditions (FC) and social influence (SI).

In numerous technology acceptance research, extrinsic motivation underlying the usage is regarded as an important factor [43]. Venkatesh et al. formulate this expectation of utility in the first UTAUT in the PE determinant [59]. A large number of technology acceptance studies have shown that PE exerts a significant influence on BI [23, 64, 65]. Extrinsic motivation exists when an act is performed to achieve a benefit that is not inherent in the action itself, while the intrinsic motivation embodied in the determinant HEDMOT is in the enjoyment of the act itself [17]. From this perspective, PE reflects the degree of extrinsic motivation or the expected outcome of use [35, 63]. Due to the functional orientation of voice assistants, the effect is also to be expected for this technology. Hence, we hypothesize:

H1: *A positive valuation of PE will positively influence the BI.*

With the use of technical systems in general and voice assistants in particular, it is crucial for acceptance that the operation is easy for the users. Otherwise, the high effort or the system-side impairment of the usage situation may constitute a barrier to adoption. The aim is therefore to achieve a positive perception of the users compared to the "degree of ease" [65], which goes hand in hand with their intentional use. Conceptually, the EE determinant considers aspects of the technology to be low in complexity and ease of use, and captures the level of self-efficacy expectation. Here, previous studies have shown that trust in one's own abilities in dealing with technical systems has a direct influence on the intention to use them [23, 65]. Amongst other aspects, for voice assistants the usability is one of the reasons for recognizing the user's requests correctly and providing adequate answers. The particular significance of the determinant EE in terms of voice assistants lies in the fact that the potential of language as an interaction medium lies above all in the simplicity and intuitive usability of the system. Hence, it can be assumed:

H2: *A positive valuation of EE will positively influence the BI.*

HEDMOT shows itself in the actual usage of voice assistants by e.g., users asking various questions to the voice assistant, expecting getting an entertaining answer. There are a number of such 'fun' features, e.g., telling fun facts or mini games, which is also reflected in the development of diverse skills [59]. These features suggest that hedonic motivation, as an incentive factor, has a positive impact on the intent to use voice assistants. This effect has been previously observed in technology research [2, 53]. When conveying information through a conversational agent, previous research demonstrated an increased level of

entertainment, which can be attributed to the mere (visual) presence of the agent [23]. People who enjoy using a technology thus seem tend to use technology more frequently and more intensely than others [16]. Thus, we propose:

H3: *With a positive hedonic motivation, the likelihood of a positive BI increases.*

As with any product also when purchasing digital voice assistants, price is a relevant factor and we thus include it into our model. The willingness to pay, however, depends on whether the price is perceived as adequate in relation to the expected benefit. The value for money is therefore positive in case the benefits of using the technology outweigh the consumer's perception of the monetary cost [65]. Technology acceptance research has shown that price value has a positive impact on behavioral intention [67, 68]. Hence, we hypothesize:

H4: *Price value has a positive impact on BI.*

A one-off or irregular use cannot be regarded as a habitual use of voice assistants; use can shift to habitualized behavior with more intense engagement. For technology acceptance research, it is important that habitual behavior expresses a strong intention to use the technology in the future. Various studies confirm habits' positive effects on behavioral intention and tendencies to adopt new technologies [39, 66]. Thus, we assume:

H5: *Habit has a positive effect on BI.*

FC refers to the resources and support required from the consumer's perspective to use the technology [64, 65, 65]. As an example, Venkatesh et al., regarding the use of mobile Internet, mention possible speed advantages that could arise from the nature of the smartphone and thus, impacting the embedded voice assistant such as Siri or Google Assistant. Consumers who have access to such resources are more likely to use the technology [65]. Thus, although some research has shown that FC is not always the best predictor, we thus propose [54]:

H6: *Facilitating conditions has a positive effect on BI.*

The construct that Venkatesh et al. [64, 65] use in UTAUT refers to how relevant caregivers prefer the use of a technology by the individual. Thus, SI means that the behavior of individuals is influenced by others because they indirectly feel pressured to fulfill the expectations that are directed to them [20]. By adapting one's own behavior, the individual intends to establish conformity with his reference group (peer group). Depending on the opinion of the peer group, the action taken may consist in the acceptance or rejection of a particular behavior [4]. Hence, we hypothesize:

H7: *Social influence will influence the BI.*

2.2. Impact of the anthropomorphism on digital voice assistants

As pointed out, all relationships discussed above are based on assumptions and theories regarding the general acceptance of technologies. Nevertheless, voice assistants differ tremendously from technologies that are not based on artificial intelligence, especially in terms of human characteristics (e.g., language). Hence, in dealing with voice assistants, naming and dialogue design do not ignore the personification of technical artefacts. Although the concept of artificial intelligence is controversial [1], the developers of the systems strive to simulate active artificial intelligence also in the interaction with voice assistants or to imitate human behavior. Since this type of design intends to generate positive effects on user perception and because of the partly critical view of this practice, we assume that this aspect can exert an influence on the acceptance of the users compared to voice assistants. That is why we specifically intend to investigate the role of anthropomorphism in this study regarding the acceptance of voice assistants.

As mentioned, anthropomorphism is defined as the tendency to attribute the actual or perceived behavior of non-human actors, human characteristics, motivations, intentions or emotions [18]. Furthermore, anthropomorphizing is based on two fundamental causes: Humans are always in search of interaction with other people and the classification of environmental influences helps to understand them and to keep them in control [58]. Therefore, even in interaction with a voice assistant, people recognize and apply patterns of behavior that they already have in everyday life to other individuals. Here, voice assistants are, for example, addressed by their name as a wake-up call, which in turn is an indicator of anthropomorphism. The design of this interaction between man and machine falls into the discipline of Human Computer Interaction (HCI), which makes dialogue a constitutive element [33]. Thus, interactive computing is characterized by the possibility of dialogue design between a computer and a machine in the form of a seamless question-and-answer behavior. This type of design of technical systems is based on natural conversations between people and intends to adapt them as much as possible. There is a clear tendency to not only mimic the structural advantages of human interaction mechanisms, but also to adopt other elements of human behavior that are not fundamentally necessary for the functionality of speech assistants, but which offer the potential "of the human user to social attributions and to trigger corresponding emotions and behaviors" [10].

According to Ortony, the creation of a personality is an important factor in contributing to the consistency of

emotional responses [49]. Indications of the positive effect of emotion-based design are provided by e.g., Becker et al. who conclude that the integration of emotions increase the credibility, liveliness and personality of an assistant [7]. In terms of affective computing, the potential of voice assistants lies in recognizing users' emotions in order to provide effective assistance [51]. Thus, emotions potentially provide an additional channel of interaction [3]. Although the development of voice assistants has not progressed far enough to establish a connection between the user's personality and the speech assistant's created personality, there is a presumption in the literature that certain matches have a positive effect on the relationship. This similarity-based perspective is also referred to as the similarity-attraction-effect, according to which people feel attracted to others who are similar to themselves [11]. According to Nass and Lee, the same effect can also be observed in terms of computer-generated voices, according to which extrovert participants prefer an extroverted voice and introverts prefer an introverted voice [46].

In general, based on previous reasoning, the most natural or realistic design of humanlike conversational agents is seen as a desirable goal in research [58]. However, the thesis of the Uncanny Valley points out that a humanlike design of robots is not always experienced as beneficial. The Uncanny Valley describes the effect that the humanlike design leads only to a certain degree to a positive increase in the perceptions of the users, to the point where the similarity is so strong that it somehow seems uncanny [44, 45]. At this stage, the design of the robot is inconsistent because, on the one hand, it is not mature enough to be congruent with a real human, but on the other hand it is already too advanced to be clearly classified as robotic. This creates the situation that the robot cannot (immediately) be assigned to a category, so that the effect of Uncanny Valley arises.

In most cases, a well-balanced anthropomorphic phenomenon, known as the "persona effect", is believed to promote the credibility and perceived usefulness and entertainment value of an agent and has a positive impact on users' attitudes to the system [38]. It can be assumed that this effect also applies to voice assistants and exerts an influence on the behavioral intent of the users. In addition to the passive attribution of human characteristics to voice assistants, the perceived anthropomorphism may also be reflected in active user actions. This phenomenon is known in research as CASA paradigm. As a result, people tend to be more responsive to computers than they would to any other person, e.g., by maintaining polite manners [47] or paying attention to presenting oneself positively [60]. In a study by Rickenberg and Reeves regarding

interactions with visual agents, users also reacted with nervousness to overly intense observation by the agent [12]. The interpretation of this behavior as a social interaction, however, meets with some critics' rejection. The users responded only to requirements that arise in the interaction. In this sense, users only stick to interpersonal interaction principles because the situation demands it. They act as if, but it never really arises the perception that it is a person [10]. Rather, the users are always aware that they communicate with or through a medium, since a complete immersion is hard to achieve. From this perspective, the psychological effect cannot be equated with interpersonal interaction [10].

However, the psychological effect of anthropomorphizing suggests that the transfer of interaction principles elicits similar associations as it is the case in interpersonal interactions. Here, perceived sociability refers to the "perceived ability of the system to perform sociable behavior" [29]. The results suggest that perceived sociability has an impact on the intent of use. Furthermore, it can be assumed that a technical system is perceived more vividly by anthropomorphization. Lifelike agents offer the potential to emotionally appeal to users [6]. To capture lifelikeness, we include animacy. Latter is based on Piaget's understanding of animation as the ability to move on its own, or to react to environmental influences [6]. In connection with technical systems it can also be applied to "artificial intelligence".

The findings suggest that voice assistants are inherent in a degree of human similarity through the use of natural language [58]. Both language as a constitutive element of human communication and interactive dialogue design thus lead to a personification of digital voice assistants. Picard points out, however, that as the system becomes more complex, the complexity of the user's requirements for the system also increases, so that voice assistants, whose design is very human-oriented, must also meet these requirements in order to be perceived positively by users [52]. First, the similarity attraction theory confirms that people are more attracted to others when certain similarities exist [11]. Second, Osgood and Tannenbaum's dissonance theory implies that people favor a congruence between themselves and the object [50], i.e., the voice assistant. Thus, if this fit is given, the assistant is perceived more positively, which in turn should lead to a higher level of likeability.

Based on prior reasoning, we thus, use animacy, perceived sociability and humanlike fit as appropriate constructs for anthropomorphism. There are several characteristics, which might influence, for instance, the perceived animacy (e.g., agreeableness [5], humanlike appearance [55] or volition [48]). However, not all cues are necessary to support a perceived animacy. In

addition, a deeper analysis of human characteristics would use dimensions of these main variables anyway. In addition, the so-called fit is an important influencing factor of "interpersonal" acceptance (e.g., celebrity endorsement [21]). Thus, if humanlike elements are perceived in the context of voice assistants, a perceived harmony of two objects (here assistant and user) should positively affect behavioral intentions.

In summary, we argue that by using these three "meta" variables we are able to identify in a more holistic approach the impact of a humanlike variables on a behavioral intention.

Finally, the likeability is appropriate to capture the positive effects of anthropomorphism [42]. Likeability could be interpreted as an affective part of the attitude towards an object. Following the three components model of attitude, this dimension contains emotions and ties to a specific object [57]. In addition, literature shows that a positive attitude in turn usually influences the behavior of users positively, concluding, that a higher level of likeability leads to a higher level of use or interaction with the given object, i.e., the voice assistant. Based on this reasoning, we hypothesize:

H8: Humanlike characteristics will positively influence the likeability of voice assistants.

H9: Likeability has a positive impact on the actual intention to use a voice assistant.

3. Method

In order to test our hypotheses, we conducted a quantitative online survey. Beforehand, we conducted a pretest to ensure a comprehensibility of the questionnaire.

In order to ensure a preexisting experience of the participants with voice assistants, i.e., on a smartphone or general smart device, we asked how familiar they are with voice assistants, how often they use them or watched someone using this technology as well as with which assistants they are dealing (Siri, Alexa, Google Assistant, Cortana, Bixby or others). Participants with no experience at all were excluded from the questionnaire. Latter was designed to only include people who have already come into contact with the use of voice assistants. Finally, 283 (average age was $M = 32.9$, 47.1% female, $SD = 12.90$) random undergraduates and participants were acquired, through links in university newsletter, in social media and in online groups.

Our operationalization of the measures is based on well-established scales of current literature. The scale for performance expectancy (3 items, e.g., "I find voice assistants useful in my daily life.", $\alpha = .93$), effort expectancy (4 items, e.g., "Learning how to use voice assistants is easy for me.", $\alpha = .89$), social influence (3

items, e.g., “People who influence my behavior think that I should use voice assistants”, $\alpha = .95$), facilitating conditions (4 items, e.g., “I have the resources necessary to use voice assistants.”, $\alpha = .85$) and behavioral intention (3 items, e.g., “I plan to use voice assistants frequently.”, $\alpha = .94$) were based on the UTAUT model by Venkatesh et al. [64]. The scale for price value (3 items, e.g., “Voice assistants are good value for the money.”, $\alpha = .89$) was adopted from Venkatesh et al. [65]. Since the survey also refers to voice assistants on smart phones, which are already pre-installed as additional functions, the assessment of the price level for those using voice assistants exclusively via the smart phone has proved to be problematic. Therefore, the option “Do not know” was added as missing value. The construct habit (4 items, e.g., “The use of voice assistants has become a habit for me.”, $\alpha = .80$) was operationalized by using a scale introduced by Limayem et al. [40]. The scale for hedonic motivation (3 items, e.g., “Using voice assistants is enjoyable.”, $\alpha = .89$) was based on the work of Kim et al. [36]. All items were measured using a 7-point Likert scale (1 = totally disagree, 7 = totally agree).

To evaluate the phenomenon of anthropomorphism, perceived sociability (4 items, e.g. “I feel the voice assistant understands me.”, $\alpha = .86$) was measured based on a scale by Heerink et al. [29]. Animacy was measured using semantic differentials, via 5 items, e.g., how “machine-like” or “humanlike” the voice assistant is ($\alpha = .85$) [6]. Likeability (5 items, e.g. “dislike” – “like”, $\alpha = .90$) was measured based on the scale of Monahan [42]. In addition, humanlike-fit between the assistant and the user (5 items, e.g. “similar” – “dissimilar”, $\alpha = .90$) [8] was examined in terms of people's attitudes to human similarity in speech assistants in general. Perceived sociability was also measured via a 7-point Likert scale, whereas animacy, humanlike-fit and likeability were collected using semantic differentials.

4. Results

In order to test the proposed hypotheses, we conducted structural equation modelling using SmartPLS. The reason for a PLS method is based on the consideration, that the research objective is prediction and theory extension. Here, the research objective is mainly confirming an established model with slight changes. The latter is more likely to be achieved by using a PLS method than by covariance-based methods [25].

The R-squared (adjusted R-squared) of the dependent variables reports a high value of .793 (.787) for “behavioral intention” and mid-sized value .394 (.387) for “likeability.” In addition, variance inflation factors (VIF) were tested, all below 4.0 and thus below the

recommended threshold of 10 [25]. Hence, we conclude that multicollinearity is not a problem at all. Finally, for all measures, the average variance extracted (AVE) was above the cutoff value of .5 [20]. In addition, the discriminant validity of the measures was tested, i.e., if a construct shares more variance with its measures than it does with other constructs in the model. Thus, the square root of the AVE exceeds the intercorrelations of the construct with any other construct in the model [22]. Table 1 reports the results of our estimation.

Table 1. Report of the results			
	Stand. Coef.	T-Statistic	VIF
Performance Expectancy → BI	.317***	4.993	3.701
Effort Expectancy → BI	-.004 ^{ns}	0.129	1.388
Hedonic Motivation → BI	.248***	5.576	2.510
Price Value → BI	.027 ^{ns}	0.767	1.604
Habit → BI	.331***	6.803	2.498
Facilitating Conditions → BI	-.038 ^{ns}	1.086	1.469
Social Influence → BI	-.007 ^{ns}	.0199	1.368
Likeability → BI	.142***	3.502	1.978
Humanlike-Fit → Likeability	.325***	5.490	1.138
Animacy → Likeability	.279***	5.011	1.544
Perceived Sociability → Likeability	.213***	3.625	1.623

Note: BI = Behavioral Intention, N = 283, PLS algorithm: maximum iterations = 300; bootstrapping procedure: cases = 283; Samples = 5000; *significant at $p < .05$, **significant at $p < .01$, ***significant at $p < .001$.

First of all, our results show that not all factors of classical technology acceptance also hold true for voice assistants. PE is highly significant and influences the BI. If the voice assistant is able to adequately fulfill his actual function, which is based on the utilitarian benefit, then it will positively influence the BI. This effect is probably due to the fact that the satisfactory feeling that comes with the achievement of the goals of use can reach high levels in its intensity which is confirmed in H1.

Surprisingly, H2 is not supported. We were therefore unable to establish that there is a connection between PE and EE [15]. Unlike previous studies [65] voice assistants might be more robust for negative effects by user perceived errors. Consciously seen as developing technology, this awareness could generally trigger a higher acceptance. Thus, some users rate the effort as low and some, depending on their common usage situation as high. Both facts could lead to this unexpected result and no clear significant effect direction.

Voice assistants are not only used as a source of information or facilitating task completion, but also because they are more likely to be considered when the interaction is fun [61], which confirms H3. If the users' interaction is enjoyable and the users perceive pleasure while interacting with the assistants, the usage intention increases.

However, we cannot observe a positive effect of price value on BI in our data. Hence, H4 cannot be confirmed. This result differs from Xu and Yang [67, 68]. One explanation might be that voice assistants are now integrated in smartphones and thus, a statement about the pure price of the voice assistant might be considered difficult. This could be interpreted as a part of technology (e.g., gadget), but not as cost related part itself. Hence, there might be a difference between the perception of the actual costs of the assistant and the costs of the carrier technology. Latter is interesting because there are standalone devices, e.g., smart speakers, where the only purpose is to offer an interface to the assistant.

Finally, hypothesis H5 can be supported. According to previous studies [39, 66] habit has a positive effect on BI. Users with habitualized behavior tend not only to evaluate new successor products, but to adapt them out of habit [40], which is an important finding especially for companies. The aim should be that users see the voice assistant as indispensable in everyday life.

As Rana et al. [54] have shown that facilitating conditions are not the strongest predictor for the BI, in our data it even has no significant influence at all on the intention to use a voice assistant. Hence, H6 is not supported. One explanation might be, that contrary to other technologies (e.g., mobile payment), no obvious counterparts (e.g., for mobile payments a NFC terminal) are needed to use the technology. So, users do not see those "resources" as critical.

Different to previous studies [4, 7, 20, 65] H7 is not supported as well. However, one might argue that voice assistants could be interpreted as second order product and as a part of a larger ecosystem. Contrary, mobile phones or services (e.g., Amazon) are first order products. Thus, users need a specific phone enabled to use a specific voice assistant [56]. Hence, the social influence might affect the actual system, e.g., someone recommends an Android phone, but not the actual intention to use a specific voice assistant. Latter is preordained by the system and thus, not directly affect by social influence.

While these results are more or less common with previous results regarding technology acceptance, in our analyses, we focused also on the impact of the voice assistants' human characteristics and the likeability on the behavioral intention.

First of all, in terms of interaction with digital voice assistants, the results show that the more positive the interaction is perceived, the higher is consumers' likability of the voice assistant. Thus, is the voice assistant perceived as active and lively, the evaluation of them increases, which could already be confirmed.

The positive impact on likeability of a perceived humanlike-fit between the assistant and the actual user is confirmed. The more similar and fitting the assistant is, the more it affects the user, the higher the likelihood is that he will like it. In summary, we can confirm H8. Moreover, the impact of likeability on BI is highly significant and positive. So, it can be postulated that a positive impression of the voice assistant has a positive effect on the use intention. Thus, H9 can be confirmed.

According to the CASA paradigm, users transfer human interactions and attributes to voice assistants, the logical consequence is that the personification is important as the study confirmed.

Based on previous reasoning it is clear that voice assistants show human attributes. The results reveal that these traits have a positive impact on how digital voice assistants are perceived. As a result, better perception leads to a greater likelihood which in turn influences the users' intention to use the technology. In summary, anthropomorphism plays a significant role and should be considered by companies as important influential factor and design element of digital assistants. In addition, it highlights some future research opportunities, such as further research on the interaction with speech-based technology.

5. General Discussion and Implications

The main purpose of the present study was to investigate anthropomorphism in the context of digital voice assistants and to further extend the Venkatesh, et al.'s UTAUT2 [65]. The results confirm most of the predicted hypotheses and substantiate to take the aspects of human characteristics for the actual user behavior in context of voice assistants into account. Regarding our research questions, our results clearly show that (1) anthropomorphism in general plays a role concerning the behavioral intention for voice assistants and thus, should be considered. Here, (2) a humanlike-fit has highest impact on a human driven likeability. By addressing a gap in literature, we show that a "conservative" view of technology is not sufficient for voice assistants, since human characteristics play a considerable role for this type of technology. Nonetheless, (3) further relevant drivers referring to the UTAUT2 model are PE, HEDMOT and habit.

The fact that voice assistants do not only follow human instructions, but also understand, learn and adequately respond to them, creates a kind of dialogue.

Based on natural language, this dialogue between man and machine should be as authentic as possible. In this regard, the results of the study have shown that it is more acceptable for voice assistants to be attributed to certain human characteristics. Unlike the Uncanny Valley paradox, we cannot confirm that humanlike characteristics will be evaluated negatively. Contrary, our results stress the importance anthropomorphism for actual BI for voice assistants.

In addition to these new findings, the study has shown that price value and effort expectancy do not influence the intention to use voice assistants. With the determinant price value, it could be that smart speakers are currently affordable and therefore acquisition costs are not so important at the moment. The follow-up costs for other devices could be rated higher. In addition, most smartphones have already integrated voice assistants and therefore a differentiated assessment of the price per se is difficult. For this purpose, a further consideration would be interesting, in which then the differentiation of price perception after stationary devices and voice assistants in mobile phones is made. Surprisingly social influence and facilitating conditions are not significant. The reason for latter could be that integrated voice assistants need no further technical counterparts since they are part of e.g., mobile phones. Even a stand-alone voice assistant only needs an internet connection which is standard nowadays or, at least, is not seen as a critical resource. With regard to the social influence, assistants might be interpreted as second order products. Here, we see the necessary device as first order product (for instance mobile phone and iOS vs. Android) [56], which is influenced by social peers, i.e., the social influence does not occur per se on the assistant, but determines which can be used.

The R squared for both variables showed that our predictors contribute significantly to the explained variance. Nevertheless, additional factors that are not covered here, such as the importance of being extensible by other devices in smart home technology, or the convenience of doing things quickly and easily should be investigated. Here, the context in which voice assistants are used could be examined to see if additional influencing factors need to be considered. Moreover, human characteristics might be interpreted and valued differently depending on the cultural background [31]. Hence, multicultural research should be sought to ensure comparability, because these human factors might be perceived differently. In addition, the features that voice assistants have so far been not available in all markets, which might impact the perceived usefulness.

Moreover, in terms of human characteristics and the related perception, future research should examine whether a male or female voice is more appropriate for the context.

Further research about voice assistants' mistakes, such as the misunderstanding of human instruction and more effort to reach the goal, could lead to interesting results in another study concerning effort expectancy.

In addition, it would be interesting to differentiate exactly which human qualities are perceived as positive and thus increase the intention and which attributes have a negative impact. Since, to our knowledge, anthropomorphism has so far received little attention in this context, we believe that we can contribute an essential part in the technology acceptance research. For future studies on language-based dialog systems and advanced artificial intelligence of non-ECAs, the UTAUT2 approach should be adapted around the confirmed determinants.

The present study also offers a broad range of implications for management. Especially in terms of technical functionality, it is important to emphasize the usefulness of the voice assistants. In terms of further networking via smart home, this field offers great potential for further expanding market shares. Once the voice assistant has been integrated into everyday life, habitualized behavior can manifest itself, making it indispensable. Latter might be beneficial for companies, because future generations and related devices are rather bought without having to be reevaluated [40, 41]. Through the impact of performance expectancy and habit, increasing application usability and customer loyalty is a primary goal in helping to increase usage. It can be assumed that with increasing benefits of the offer, a foundation will be created that will favor habitualized usage scenarios.

Previous literature as well as the results of the study confirm that it is reasonable for companies not only to see the voice assistant as an utilitarian object in terms of usefulness. Developers should continue to make sure that the interaction is fun and enjoyable. Therefore, we recommend to develop the hedonistic character with interactive games or small features that will delight even larger groups. This could enhance the humanity of voice assistants, as humor can be identified as a human trait. The results have shown that the humanization of voice assistants is a successful driver in terms of usage intention, which should support our argumentation here.

We also recommend creating voice assistants in a form that users like to communicate with. Courtesy forms such as they exist in conversations between people, can increase the sympathy, create a positive image and ensure that the assistant is perceived as a pleasant conversation partner, which will be reflected in the end in a positive intention to use. In addition, we show that the voice assistant is perceived to be more positive if it is more similar to the users and gives a coherent overall picture. Here we argue that it is important for companies to attribute common positive

human character traits to the voice assistant and have them recalled in the appropriate situations. For the future, it would certainly be interesting to see that the digital voice assistant is equipped with different characters and adapts to the appropriate user based on various criteria in order to be perceived as similar or complementary to the user himself.

However, our study also has its limitations. First, we only considered persons who are frequently in contact with digital voice assistants or at least watched someone consciously using this technology. Primarily, the implications can be only transferred to all people with knowledge about this technology. Thus, users with less knowledge should be considered in a further study. Second, the sample was conducted among consumers in Germany and contains primarily students. Prior research shows several important differences in personalities with respect to technology adoption (e.g., data privacy). Latter might also impact the perception of humanlike elements and weighting their importance. Thus, further research should address these issues.

6. References

- [1] <http://www-formal.stanford.edu/jmc/history/>, accessed 6-15-2018.
- [2] Alalwan, A.A., Y.K. Dwivedi, and N.P. Rana, "Factors influencing adoption of mobile banking by Jordanian bank customers - Extending UTAUT2 with trust", *International Journal of Information Management*, 37(3), 2017, pp. 99–110.
- [3] Ball, E., "A Bayesian heart Computer recognition and simulation of emotion", in *Emotions in Humans and Artifacts*.
- [4] Barth, S. and M.D.T.d. Jong, "The privacy paradox – Investigating discrepancies between expressed privacy concerns and actual online behavior – A systematic literature review", *Telematics and Informatics*, 34(7), 2017, pp. 1038–1058.
- [5] Bartneck, C., M.v.d. Hoek, O. Mubin, and A.A. Mahmud, "'Daisy, Daisy, give me your answer do!': Switching off a robot", *Proceedings of the Second ACM SIGCHI/SIGART Conference on Human-Robot Interaction*, 2007, pp. 217–222.
- [6] Bartneck, C., C. Kulic, E. Croft, and S. Zoghbi, "Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots", *International Journal of Social Robotics*, 1(1), 2009, pp. 77–81.
- [7] Becker, C., S. Kopp, and I. Wachsmuth, "Why emotions should be integrated into conversational agents", 2007, pp. 49–68.
- [8] Becker-Olsen, K.L. and R.P. Hill, "The Impact of Sponsor Fit on Brand Equity - The Case of Nonprofit Service Providers", *Journal of Service Research*, 9(1), 2006, pp. 73–83.
- [9] Bellegarda, J.R., "Large-Scale Personal Assistant Technology Deployment – the Siri Experience", *Interspeech – Proceedings of the Annual Conference of the International Speech Communication Association*, 2013, pp. 2029–2033.
- [10] Bente, G., N. Krämer, and A. Petersen, *Virtuelle Realitäten*, Göttingen, 2003.
- [11] Bernier, E.P. and B. Scassellati, "The Similarity-Attraction Effect in Human-Robot Interaction", *IEEE 9TH International Conference on Development and Learning*, 2010, pp. 286–290.
- [12] Brahmam, S. and A. de Angeli, "Gender affordances of conversational agents", *Interacting with Computers*, 24(3), 2012, pp. 139–153.
- [13] Callejas, Z., R. López-Cózar, N. Abalos, and N. Griol, "Affective Conversational Agents – the Role of Personality and Emotion in Spoken Interactions".
- [14] Cassell, J., C. Pelachaud, N.I. Badler, M. Steedman, B. Achorn, T. Becket, B. Douville, S. Prevost, and M. Stone, "Animated Conversation – Rule-Based Generation of Facial Expression, Gesture & Spoken Intonation for Multiple Conversational Agents", *SIGGRAPH '94 Proceedings of the 21st annual conference on Computer graphics and interactive techniques*, 1994, pp. 413–420.
- [15] <https://www.capgemini.com/de-de/wp-content/uploads/sites/5/2018/01/dti-conversational-commerce-3.pdf>, accessed 3-15-2018.
- [16] Davis, F.D., R.P. Bagozzi, and P.R. Warshaw, "Extrinsic and Intrinsic Motivation to Use Computers in Workplace", *Journal of Applied and Social Psychology*, 22(14), 1992, pp. 1111–1132.
- [17] Deci, E.L., "The Effects of Contingent and Noncontingent Rewards and Controls on Intrinsic Motivation", *Organizational Behavior & Human Performance*, 8(2), 1972, pp. 217–229.
- [18] Epley, N., A. Waytz, and J.T. Cacioppo, "On seeing human a three-factor theory of Anthropomorphism", *Psychological Review*, 114(4), 2007, pp. 864–886.
- [19] Ferrara, E., O. Varol, C. Davis, F. Menczer, and A. Flammini, "The Rise of Social Bots", *Communications of the ACM*, 59(7), 2016, pp. 96–104.
- [20] Festinger, L., "Some Attitudinal Consequences of Forced Decisions", *Acta Psychologica*, 15, 1959, pp. 389–390.
- [21] Fleck, N., M. Korchia, and I. Le Roy, "Celebrities in advertising: Looking for congruence or likability?", *Psychology & marketing*, 29(9), 2012, pp. 651–662.
- [22] Fornell, C. and Larcker D.F., "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error", *Journal of Marketing Research*, 18(1), 1981, pp. 39–50.
- [23] Fridin, M. and M. Belokopytov, "Acceptance of socially assistive humanoid robot by preschool and elementary school teachers", *Computers in Human Behavior*, 33(4), 2014, pp. 23–31.
- [24] Goudey, A. and G. Bonnin, "Must smart objects look human? Study of the impact of anthropomorphism on the acceptance of companion robots", *Recherche et Applications en Marketing*, 31(2), 2016, pp. 2–20.
- [25] Hair, J.F., C.M. Ringle, and M. Sarstedt, "PLS-SEM: Indeed a Silver Bullet.", *Journal of Marketing Theory and Practice*, 19(2), 2011, pp. 139–151.
- [26] Harris, R.A., *Voice interaction design: Crafting the new conversational speech systems*, Morgan Kaufmann Publishers, San Francisco, 2005.
- [27] Hauswald, J., M.A. Laurenzano, Y. Zhan, C. Li, A. Rovinski, A. Khurana, R.G. Dreslinski, T. Mudge, V. Petrucci, L. Tang, and J. Mars, "An Open End-to-End Voice and Vision Personal Assistant and Its Implications for Future Warehouse Scale Computers", *ACM SIGPLAN Notices*, 50(4), 2015, pp. 223–238.
- [28] Hayes-Roth, B., "An architecture for adaptive intelligent systems", *Artificial Intelligence*, 72(1-2), 1995, pp. 329–365.
- [29] Heerink, M., B. Kröse, V. Evers, and B. Wielinga, "Assessing Acceptance of Assistive Social Agent Technology by Older Adults – the Almere Model", *International Journal of Social Robotics*, 2(4), 2010, pp. 361–375.

- [30] Horstmann, A.C., N. Bock, E. Linhuber, J.M. Szczuka, C. Straßmann, and N.C. Krämer, "Do a robot's social skills and its objection discourage interactants from switching the robot off?", *PLoS ONE*, 13(7), 2018.
- [31] House, R.J., ed., *Culture, leadership, and organizations: The GLOBE study of 62 societies*, Sage Publ, Thousand Oaks, California, 2004.
- [32] <https://www.theguardian.com/technology/2018/jan/06/how-smart-speakers-stole-the-show-from-smartphones>, accessed 3-19-2018.
- [33] Imaz, M. and D. Benyon, *Designing with Blends Conceptual Foundations of Human-Computer Interaction and Software Engineering*, Cambridge, 2007.
- [34] <https://www.campaignlive.co.uk/article/just-say-it-future-search-voice-personal-digital-assistants/1392459>.
- [35] Kim, H.-W., H.C. Chan, and S. Gupta, "Value-based Adoption of Mobile Internet – An empirical investigation", *Decision Support Systems*, 43(1), 2005, pp. 111–126.
- [36] Kim, S.S., N.K. Malhotra, and S. Narasimhan, "Two Competing Perspectives on Automatic Use – A Theoretical and Empirical Comparison", *Information Systems Research*, 16(4), 2005, pp. 418–432.
- [37] Kopp, S., L. Gesellensetter, N.C. Krämer, and Wachsmuth I., "A conversational agent as museum guide: design and evaluation of a real-world application", in *Intelligent Virtual Agents*.
- [38] Lester, J.C., S.A. Converse, S.E. Kahler, S.T. Barlow, B.A. Stones, and R.S. Bhogal, "The persona effect Affective impact of animated pedagogical agents", *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*, 1997, 359-366.
- [39] Lewis, C.C., C.E. Fretwell, J. Ryan, and J.B. Parham, "Faculty Use of Established and Emerging Technologies in Higher Education - A Unified Theory of Acceptance and Use of Technology Perspective", *International Journal of Higher Education*, 2(2), 2013, pp. 22–34.
- [40] Limayem, M., S.G. Hirt, and C.M.K. Cheung, "How habit limits the predictive power of intention: The case of information systems continuance", *MIS Quarterly*, 31(4), 2007, pp. 705–737.
- [41] Magotra, I., J. Sharma, and S.K. Sharma, "Assessing personal disposition of individuals towards technology adoption", *Future Business Journal*, 2(1), 2016, pp. 81–101.
- [42] Monahan, J.L., "I don't Know it but I like you - the influence of non-conscious affect on person perception", *Human Communication Research*, 24(4), 1998, pp. 480–500.
- [43] Moon, J.-W. and Y.-G. Kim, "Extending the TAM for a World-Wide-Web context", *Information & Management*, 38(4), 2001, p. 217.
- [44] Mori, M., "The Uncanny Valley", *Energy*, 7(4), 1970, pp. 33–35.
- [45] Mori, M., "The Uncanny Valley", *IEEE Robotics & Automation Magazine*, 19(2), 2012, pp. 99–100.
- [46] Nass, C. and K.M. Lee, "Does Computer-Synthesized Speech Manifest Personality? Experimental Tests of Recognition, Similarity-Attraction, and Consistency-Attraction", *Journal of Experimental Psychology*, 7(3), 2001, pp. 171–181.
- [47] Nass, C., J. Steuer, and E.R. Tauber, "Computers are Social Actors", *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1994, pp. 72–78.
- [48] Nigram, M.K. and D. Klahr, "If robots make choices, are they alive? Children's judgments of the animacy of intelligent artifacts", *Proceedings of the 22nd Annual Conference of the Cognitive Science Society*, 2000.
- [49] Ortony, A., "On making believable emotional agents believable".
- [50] Osgood, C.E. and P.H. Tannenbaum, "The principle of congruity in the prediction of attitude change", *Psychological Review*, 62(1), 1955, pp. 42–55.
- [51] Picard, R.W., *Affective Computing*, Cambridge, 1997.
- [52] Picard, R.W., "Affective computing", *International Journal of Human-Computer Studies*, 59(1-2), 2003, pp. 55–64.
- [53] Raman, A. and Y. Don, "Preservice Teachers' Acceptance of Learning Management Software - An Application of the UTAUT2 Model", *International Educational Studies*, 6(7), 2013, pp. 157–164.
- [54] Rana, N.P., M.D. Williams, Y.K. Dwivedi, and J. Williams, "Theories and Theoretical Models for Examining the Adoption of E-Government Services", *e-Service Journal*, 8(2), 2012, pp. 26–56.
- [55] Riek, L.D., T.-C. Rabinowitch, B. Chakrabarti, and P. Robinson, "How anthropomorphism affects empathy toward robots", *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction*, 2009, pp. 245–246.
- [56] Rochet, J.-C. and J. Tirole, "Platform Competition in Two-Sided Markets", *Journal of the European Economic Association*, 1(4), 2003, pp. 990–1029.
- [57] Rosenberg, M.J., *Attitude organization and change: An analysis of consistency among attitude components*, Yale Univ. Press, New Haven, Conn., 1960.
- [58] Seeger, A.-M., J. Pfeiffer, and A. Heinzl, "When Do We Need a Human? Anthropomorphic Design and Trustworthiness of Conversational Agents", *Proceedings of the Sixteenth Annual Pre-ICIS Workshop on HCI Research in MIS*, 15(1-6), 2017.
- [59] <https://www.amazon.de>, accessed 5-27-2018.
- [60] Sproul, L., M.R. Subramani, and S. Kiesler, "When the Interface Is a Face", *Human-Computer-Interaction*, 11(2), 1996, pp. 97–124.
- [61] Suzuki, N., K. Ishii, and M. Okada, "Talking Eye - autonomous creature as accomplice for Human", *Proceedings of the 3rd Asia Pacific Conference on Computer Human Interaction*, 1998, pp. 409–414.
- [62] Tadeusiewicz, R., "Speech in Human System Interaction", *3rd Conference on Human System Interactions*, 2010, pp. 2–13.
- [63] Venkatesh, V., "Creation of favorable user perceptions: Exploring the role of intrinsic motivation", *MIS Quarterly*, 23(2), 1999, pp. 239–260.
- [64] Venkatesh, V., M.G. Morris, G.B. Davis, and F.D. Davis, "User acceptance of information technology: Toward a unified view", *MIS Quarterly*, 27(3), 2003, pp. 425–478.
- [65] Venkatesh, V., J.Y.L. Thong, and X. Xu, "Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology", *Management information systems : mis quarterly*, 36(1), 2012, pp. 157–178.
- [66] Wong, C.H., G.W.H. Tan, S.P. Loke, and K.B. Ooi, "Mobile TV: A new form of entertainment?", *Industrial Management & Data Systems*, 114(7), 2014, pp. 1050–1067.
- [67] Xu, X., "Understanding Users' Continued Use of Online Games - An Application of UTAUT2 in Social Network Games", *Sixth International Conferences on Advances in Multimedia*, 2014, pp. 58–65.
- [68] Yang, S., "Understanding Undergraduate Students' Adoption of Mobile Learning Model - A Perspective of the Extended UTAUT2", *Journal of Convergence Information Technology*, 8(10), 2013, pp. 969–979.