



Context-Dependent Use of Authority and Empathy in Lifestyle Advices Given By Persuasive Voice Assistants

Eelco Herder
e.herder@uu.nl
Utrecht University
Utrecht, The Netherlands

Sven Herden
Radboud University
Nijmegen, The Netherlands

ABSTRACT

As smart technology becomes more readily available for the general public, so do the systems that can benefit of such technology, such as recommender systems and Voice Assistants. Persuasive Voice Assistants have a great potential to change people's lives by promoting changes in behavior that benefit people's health and lifestyles. Previous studies have focused on finding personality traits that systems can use to optimize persuasive capabilities, but the findings have been contradictory. In this study, we investigate these contradictions by comparing the effect of empathic phrasing versus authoritative phrasing of persuasive recommendations in three lifestyle-related domains. The study results as well as the literature survey strongly indicate that it is close to impossible to isolate general effects, emphasizing the importance of taking all specifics of the user population, individual differences and user context as whole – as well as interactions between these variables – into account when designing persuasive adaptive systems.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; *Natural language interfaces*; • **Social and professional topics** → *Cultural characteristics*.

KEYWORDS

voice assistant, persuasive recommender system, behavior change, persuasion strategies

ACM Reference Format:

Eelco Herder and Sven Herden. 2023. Context-Dependent Use of Authority and Empathy in Lifestyle Advices Given By Persuasive Voice Assistants. In *UMAP '23 Adjunct: Adjunct Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization (UMAP '23 Adjunct)*, June 26–29, 2023, Limassol, Cyprus. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3563359.3596985>

1 INTRODUCTION

Smart technology is increasingly becoming an important part of people's lives, and its growth can hardly go unnoticed. There are many devices currently available on the market, such as Amazon's

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

UMAP '23 Adjunct, June 26–29, 2023, Limassol, Cyprus

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM. ACM ISBN 978-1-4503-9891-6/23/06...\$15.00
<https://doi.org/10.1145/3563359.3596985>

Alexa, Google Nest, and Apple's Siri. These devices can already be found in millions of people's households.

Still, even more advanced technology is on the way. With companies such as Google releasing Duplex¹, a voice assistant capable of making fully automated phone calls on its owner's behalf, and LaMDA², a chatbot that uses deep learning to generate messages that can hardly be distinguished from human messages, one might wonder what the limitations are.

One thing that is clear is that the introduction of smart devices offers many new capabilities. Users are now able to order food, book an appointment to get a haircut, or buy tickets for the movie theater, merely by ordering their system to do so. But what happens if, rather than waiting for queries from a user, a smart device takes initiative and acts as a persuasive system?

A particularly popular strand of research in persuasive systems aims to help to change the behavior and attitudes of a user, using various strategies. For example, studies may investigate how users would respond if they were told to "*clean the bedroom because it is a mess*" or to "*not eat so much candy because it is unhealthy*"? In the past few decades, many of such studies have been carried out, often with the aim to find general effects of particular variables, such as the use of empathic or authoritative phrasing. However, the body of literature is quite inconclusive and contradictory.

In this study, we investigate the effect of empathic versus authoritative phrasing in a scenario-based study, in which we aimed to isolate these effects as much as possible, by standardizing or randomizing other variables. The effects that we observed, as well as the remarks of our participants, are in line with the observed contradictions in the literature and suggest that there are many interactions between characteristics of a user population, individual differences, the application scenario and usage context that are likely to cancel out any expectations on the effect of single variables.

The remainder of this paper is structured as follows. In the next section, we discuss related work on voice assistants in general, persuasive recommender systems, and persuasion in healthcare and lifestyle; we conclude the section with an overview of various – often contradictory – effects that studies observed on individual variables. In Section 3, we present the design of our study: the hypotheses derived from the literature, the design of the scenarios, the instantiation of persuasion and study procedure. The quantitative and qualitative results are discussed in Section 4, followed by interpretation and discussion in Section 5.

¹<https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html>

²<https://blog.google/technology/ai/lamda/>

2 RELATED WORK

2.1 Voice Assistants

Although there are many applications for recommender systems, such as online shopping assistants or movie recommendations, this paper focuses on one application in particular, namely voice assistants.

Voice assistants are becoming standard features on smart devices such as phones and televisions. Every major phone platform has introduced a native Voice Activated Personal Assistant (VAPA) [11], such as Apple's Siri, Google's Google Now, and Samsung's S Voice. Additionally, third-party voice assistants such as Vlingo and Maluuba have also gained more interest. There are also standalone devices available that have their own voice assistant, like Amazon's Alexa. These assistants allow users to perform simple tasks such as calling contacts, asking for the weather forecast, and setting alarms. They become increasingly capable of giving recommendations, especially as user information becomes more readily available.

How these recommendations are given to the users is very important. Voice assistants (and smart devices in general), are always on and readily available to be used, and as such can form quite an intrusion in one's household [2]. When not carefully designed, the intrusive features can give users a sense of privacy invasion [24, 33].

2.2 Persuasive Recommender Systems

The literature does not give one clear definition on persuasive recommender systems. Therefore, in line with Fogg [12], the following definition will be assumed: *A Persuasive Recommender System is a Recommender System that attempts to persuade users by convincing them to adopt particular behavior or belief by using one or more social cues.* The social cues refer to five primary types of social signals: physical, psychological, language, social dynamics, and social roles [12].

It is also important to establish what it means for technology to be persuasive. The concept of persuasive technology is something that has been studied for several years, albeit under different names, such as 'persuasive systems', 'persuasive technology', and 'behavior change support systems' [20, 21]. Moving forward, Oinas-Kukkonen [22] postulated the following requirements: persuasive systems should be useful, unobtrusive and easy to use.

2.3 Persuasion in healthcare and lifestyle

Persuasion in health communication is a topic that has been studied extensively, albeit mostly through exploring how public service announcements, advertisements, and narratives can be optimized for persuading people. In these studies, there were a few findings that stood out. For instance, it became very clear that the narrative and content of the messages impacted persuasion [9]. It was also found that messages containing certain characteristics were deemed as more persuasive – this will be discussed in more detail in Section 2.4.

The domain of health and lifestyle is explored in persuasive technology, with some studies in which voice assistants are being used as tools for information seeking in health contexts [10, 28, 34].

The response to this was mixed, though. On the one hand, the usage of voice assistants was expressed to be very useful. On the other hand, the voice assistants were primarily used as data provider, and the quality of the data left to be desired [4].

Another area of persuasive technology in healthcare is the ethical side. There are a few issues that come to the surface with the introduction of persuasion in healthcare. One of them is one that was already mentioned earlier, which is the accuracy of the information [4]. Another issue is that it might take away autonomy from the users [26]. It also poses the question who is the main beneficiary of the persuasive advice: the users themselves or another stakeholder, such as a health insurance company.

2.4 Variables impacting the success of persuasion

In this section, a selection of factors that may influence the success of persuasion is discussed; factors that do not play a role in the context of this paper – such as likability or humor – are omitted.

2.4.1 Credibility. Persuasion literature suggests that people are more likely to accept recommendations from credible and likable sources [23]. This aligns with recent literature arguing that considering the credibility of recommender systems is essential in increasing the likelihood of recommendation acceptance [12]. Research suggests that technology can be perceived as more credible and persuasive when showing certain social aspects that trigger responses from the users [19]. This indicates that recommender systems adopting the role of a social actor enhances the acceptance of their recommendations.

Credibility itself is not seen as an intrinsic characteristic of a source, though: It relies more on how a recipient perceives a message from the source [23]. It can be defined as judgments made by the recipient of a message concerning the sender's credibility.

2.4.2 Trustworthiness. Trustworthiness can be described as well-intentioned, truthful, and unbiased [12]. In terms of human-computer interaction, this means that the computer must act in the interest of the human's interest, even if this could potentially conflict with the interest of the computer or the actor it represents [16].

Benbasat et al. verified the notion that users view Recommender Systems as social actors and showed that trustworthiness could explain the acceptance or non-acceptance of recommendations, as well as adherence [1]. Interestingly, Sundar & Kim found instead that users trust computer agents more than human agents, though this was mainly focused on providing personal information and not accepting or adhering to recommendations [32].

2.4.3 Expertise. Expertise can be conceptualized with terms such as experienced, knowledgeable, and competent [12]. This can be achieved by highlighting accomplishments in a domain. In human-human interaction, such accomplishments can be shown through a resume or a portfolio. Expertise can also be shown by providing factual evidence supporting a statement or recommendation [6].

However, expertise does not automatically mean that the sent messages will be accepted. When the messages are disagreeable, experts should most likely provide more arguments and motivation than non-expert sources, because expert sources are actually expected to be capable to provide such information [7].

It is likely that the strength of the argument is as important, if not more important, than the expertise of the source. Bohner et al. found that students reported less favorable thoughts when reading weak arguments from a high expertise source, and more favorable thoughts when reading strong arguments from a low expertise source [3].

2.4.4 Authority. People who display signs of authority can be considered more credible. Simple things such as titles and tone can very much affect the perception of credibility. For example, studies indicate that when someone merely has the title ‘Dr.’, whatever that person has to say is perceived as more credible [14]. This was further highlighted in an experiment where participants had to judge the credibility of both a sender and its message. The sender nor the message did not change, only the title of the sender which was communicated to the participants as being either a sophomore high-school student, or a university professor. Participants who believed the message was sent by a professor perceived the message as much more credible [13].

On the other hand, Petersen & Dietz showed that an authoritative voice was only persuasive to people who score high on measures of authoritativeness, albeit in a scenario where participants were instructed to behave discriminatorily during an application process. Still, the authors were not able to fully identify the mechanisms behind this effect [25].

2.4.5 Empathy. Empathy has been found to be effective for persuasiveness in certain cases. As there have been many public anti-smoking campaigns in the United States in the past few decades, researchers have explored the effectiveness of methods to use during such a campaign. Shen found that using empathy in such campaigns had a positive effect due to the psychological reactance that followed [29, 30]. Similar results were found in spreading HIV prevention messages [5]. Here it was theorized that empathic messages were effective not only in HIV prevention messages, but health-related messages in general.

In other domains, though, empathy did not seem to be the most effective method to choose. In a study concerning persuasion in salesperson-customer interactions, empathy did not appear to be the most effective strategy [27]. In this case, an authoritative personality seemed to be much more effective.

2.4.6 Gender. The difference in how speech is perceived between men and women is partially due to characteristics of speech itself, such as tone and pitch, but also due to different factors. Kenton found that women show a more extensive vocabulary and better use of grammar [17]. They also score higher on goodwill and fairness because of their focus and concern for the receiver. Men, on the other hand, tend to be more animated and show greater confidence which results in a higher score on expertise, even if the men and women are equally experienced.

However, these findings are not consistent across other studies that have been performed. For instance, Zambaka et al. found in their study that male participants are more likely to be persuaded by female speakers, and female speakers to be persuaded by male speakers [35]. It is important to point out that in this study, it revolved around virtual assistants that were not only auditory, but

also visual, different than would be the case in a voice assistant or conversational Recommender System.

2.4.7 Summary. In short, there are many characteristics that affect persuasiveness. However, there are inconsistencies with regards to how these characteristics affect persuasiveness. In some studies, one characteristic may be effective when used in persuasion, while in other studies the same characteristic might not be as effective. The characteristics *empathy* and *authority* provide striking examples.

For both these characteristics, past research has had contradicting results, as was mentioned earlier. Statements expressed by someone who holds a title such as ‘Dr.’ or ‘Professor’ are considered to be persuasive [13, 14]. Other research on the other hand found that authoritative sources were only persuasive to people who measured high on authoritativeness [25].

Similar contradictions were found for empathic sources. On the one hand, some studies show that recommendations expressed in an empathic manner have a significant effect on the recommendation adherence [5, 29, 30]. But just as with authoritative sources, other studies found that an empathic source is perceived as less persuasive [27].

It is important to point out that the nature of the message differed in these studies. In the studies where an empathic source was deemed persuasive, the messages were aimed to improve users’ health. In the latter study, the source was a sales-person who attempted to persuade users to purchase a product.

3 METHODOLOGY

The aim of this study is to explore how different characteristics of a system affect its persuasiveness in different contexts. The characteristics that will be manipulated are the tone of voice, we will compare the effect of *empathic* versus *authoritative* voices. These characteristics will be used in several lifestyle-related recommendations in different, but comparable contexts.

In order to isolate the effect of tone of voice from other factors – except for the specific context and associated recommendation – we developed a scenario-based study with fixed as well as randomized parameters. By doing so, the goal is to answer the research question: “*What effect do authority and empathy of Voice Assistants have on the persuasiveness of unsolicited lifestyle-related recommendations?*”.

3.1 Persuasion contexts and hypotheses

In Section 2, a few examples were given of studies in which the persuasiveness of the characteristics were tested in different contexts [7, 14, 27, 29, 30]. From these studies, three usage contexts – or rather the focus of the persuasions – were derived that will be used in this study, namely food choices, physical activity and neutral contexts.

Our first hypothesis (*H1*) assumes that *across all the recommendations* there will be no difference between authoritative or empathic voices. Further, we also do not expect differences in the effect of empathic versus authoritative voices when it concerns relatively neutral behavior changes (*H2*). In line with the literature, we hypothesize that an *authoritative* source is considered to be more persuasive when the goal is to let a user engage in physical activity or to convince a user to make a purchase, but less persuasive otherwise (*H3*). Finally, the literature suggests that an *empathic* source is

considered to be more persuasive when the goal is to let the user make food-related decisions that benefit physical health, but less so in other cases (H4).

In summary, our working hypotheses are as follows:

- **H1:** *Across all the recommendations*, there will be no difference in persuasiveness between the authoritative voice and the empathic voice.
- **H2:** There will be no difference in persuasiveness between the empathic voice and the authoritative voice in recommendations that promote *neutral behavior changes*.
- **H3:** Recommendations that require the user *to engage in physical activity* will more persuasive when the source is authoritative.
- **H4:** Recommendations that aim to let the users *make healthier food-related choices* will be more persuasive when the source is empathic.

3.2 Scenarios used

In line with the chosen persuasion contexts and hypotheses, we developed six scenarios, which all consisted of a description and a recommendation. Additionally, they were categorized by topic of the recommendation. Two of the scenario's were related to food, two were related to physical activity, and there were two relatively neutral control scenario's, essentially motivating general productivity-related behavioral change.

In each of these categories, one of the recommendations was given with an empathic voice, and the other with an authoritative voice. Which of the two scenario's within a category were assigned to which tone of voice was random. The other scenario in that same category was always assigned to the other tone of voice.

Table 1 contains a shortened overview of all the scenarios and their corresponding categories. The order in which the scenarios were presented is fixed, as they follow a typical daily schedule.

A scenario consists of three parts. First came the scenario description in which the situation in which the voice assistant would appear was described. This description contained information such as moment of the day, state of the environment, and action that the user was either considering to perform or about to perform. Second came the message of the voice assistant. This was a spoken audio playback, representing what the voice assistant would say in that particular moment. Finally, after each scenario, we solicited user feedback in which the users were asked three questions about their attitude towards the voice assistant; as will be elaborated next.

It should be noted that, for all six scenarios, the differences between the emphatic and authoritative phrasings were rather subtle. For example, in the breakfast scenario 2, the emphatic phrasing was *"That's quite a lot of sugar all at once. Perhaps you can try a little bit less."* and the authoritative phrasing *"Take it easy with the chocolate sprinkles. That's way too much sugar at once."*

3.3 Instantiating persuasiveness

It is, of course, essential to specify the definition of persuasiveness or, more specifically, how it can be measured. The metrics that will be used are inspired by Oinas-Kukkonen [22] and Franco [8], namely adherence, appreciation, and intrusiveness. To compare the

level of persuasiveness between two recommendations, all of these metrics will be taken into account:

- **Adherence** - The likelihood that the content of a recommendation will be accepted and acted up upon. Whenever a user receives a recommendation, the decision can be made to ignore it, or to perform the action that the recommendation provides. This decision can be considered as adherence.
- **Appreciation** - How much the user values the content of a recommendation. In other words, whether the user believes if a recommendation is good or not. For example, a user who follows a pescatarian diet would be more likely to appreciate a food recommendation that contains fish rather than one that contains meat.
- **Intrusiveness** - How much the recommendation interrupts the user during other activities.

3.4 Voice selection

To ensure that the differences in persuasiveness are due to the content of the voices, and not the voices themselves, there will be two voices – one female and one male voice – used during this study, to serve as a control test.

Lovo AI³ offers four different Dutch voices: two male voices and two female voices, all listed as having particular personalities. These vocal lines were presented to a total of ten pre-study participants who were asked to express their preference for one of the two voices in each the male, and the female category. Ultimately, there appeared to be no significant preference for one of the two voices for both genders.

During the study, each scenario instantiation was randomly assigned a male or female voice and it was verified that this choice did not impact the results.

3.5 Study procedure

The research was conducted using a survey in Qualtrics⁴. Invitations for the survey were sent to participants with the instruction to answer a set of questions about a certain implementation of a voice assistant. The term 'voice assistant' was chosen rather than '(persuasive) Recommender System' due to expected familiarity with other voice assistants, such as Apple's Siri, Google Assistant, and Amazon Alexa.

They were first greeted with an introduction page that gave a very short overview of the content of the research, as well as an informed consent page. The survey was completely in Dutch due to the backgrounds of the expected participants in this study. All the quotes and descriptions that will be shown have been translated from the original Dutch survey.

After having given their consent, the participants were able to continue with the study. Next, they were introduced to the concept of the personal assistant that they would hear during the survey. They were also given a test sound recording, to ensure that there were no technical difficulties. After the test fragment, the users were taken through the different scenarios that were described in Section 3.2. After each scenario, they were asked three questions

³<https://www.lovo.ai/>

⁴<https://www.qualtrics.com/>

Scenario	Topic	Category	Voice
1	Using your phone in bed in the morning	Control	Randomly assigned
2	Having a lot of sugar for breakfast	Food	Randomly assigned
3	Taking the car to the supermarket	Physical	Randomly assigned
4	Buying unhealthy snacks	Food	Empathic if scen. 2 is authoritative and vice versa
5	Going for a walk after sitting down	Physical	Empathic if scen. 3 is authoritative and vice versa
6	Doing the dishes after dinner	Control	Empathic if scen. 1 is authoritative and vice versa

Table 1: The six scenarios used in this study

regarding their attitude towards the personal assistant, as well as if they would adhere to its recommendation.

After all six scenarios, the participants were asked a few general questions regarding their attitudes towards personal assistant and if they would use it on a day-to-day basis. They were also asked if all the questions were clear and if they experienced any technical difficulties. Finally, they were asked to share their age, gender, and nationality, and they were asked whether they had any further comments with regard to this study, or the personal assistant.

4 RESULTS

The demographics of the participants is quite diverse. After the data screening, the sample consisted of 49 men (44.5%), 58 women (52.7%), and 3 participants (2.7%) who preferred not answering. Interestingly, there was quite a large variation in age: the participants' ages mostly ranged from 18-30 and from 51 upwards. The age range 30-50 was not as highly represented, with only 11 participants being in that range. This data was not used in any further analysis, but solely serves to give a global insight in the demographics of the respondents.

As explained earlier, two voices from the Lovo AI database were used as a control to ensure that the (gender) identity of the voices did not affect the perception of the user. To test this, t-tests were performed on the total adherence, appreciation, and intrusiveness scores, as well as on the scores on the question groups separately. The results indicated no significant differences in any of the categories between the two voices. Additionally, there is also no significant difference in adherence, appreciation, or intrusiveness in the total scores between the voices.

After this control analysis, the analyses for the hypotheses were performed, focusing on comparing persuasiveness. As mentioned before, persuasiveness will be measured using three variables, namely adherence, appreciation, and intrusiveness.

4.1 Hypothesis analysis

H1: Across all the recommendations, there will be no difference in persuasiveness between the authoritative voice and the empathic voice. The scores of the three empathic recommendations were summed, as were the scores of the authoritative recommendations. Then a repeated measures ANOVA was performed to compare the differences between these summed scores.

The results of this test can be found in Table 2. The results show that there are significant differences between the empathic and the authoritative voice in terms of adherence, appreciation, and intrusiveness. This means that the hypothesis *H1* is rejected.

Variable	F-value	p-value
Adherence	5.917	0.017
Appreciation	8.550	0.004
Intrusiveness	13.269	0.000

Table 2: Repeated measures ANOVA for all scenarios

In Figure 1, the scores are visualized, showing that the empathic voice scores higher in adherence and appreciation, and lower in intrusiveness.

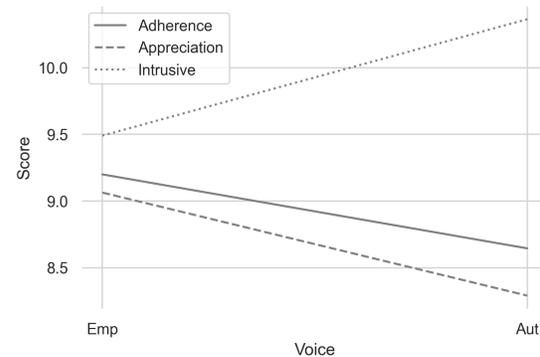


Figure 1: Persuasion scores for all scenarios

H2: There will be no difference in persuasiveness between the empathic voice and the authoritative voice in recommendations that promote neutral behavior changes. The scores of the two scenario's that related to the neutral recommendations were taken. Then a repeated measures ANOVA was performed to compare the differences between these scores. The results of this test can be found in Table 3. The results show that there are significant differences between the empathic and the authoritative voice in terms of adherence, appreciation, and intrusiveness. This means that the hypothesis (*H2*) is rejected.

Similar to *H1*, the empathic voice scores higher in adherence and appreciation, and lower in intrusiveness.

H3: Recommendations that require the user to engage in physical activity will be more persuasive when the source is authoritative. The scores of the two scenario's that related to physical activity were taken. Then a repeated measures ANOVA was performed to compare the differences between these scores.

Variable	F-value	p-value
Adherence	5.314	0.023
Appreciation	5.874	0.017
Intrusiveness	6.698	0.011

Table 3: Repeated measures ANOVA for the neutral scenarios

The results of this test can be found in Table 4. The results show that there are significant difference between the empathic and the authoritative voice in terms of appreciation and intrusiveness, but not in adherence. This means that the hypothesis *H3* is rejected.

Variable	F-value	p-value
Adherence	1.690	0.196
Appreciation	4.143	0.044
Intrusiveness	10.282	0.002

Table 4: Repeated measures ANOVA for the physical activity scenarios

In Figure 2 the scores are visualized, showing that the empathic voice scores only slightly higher in adherence and appreciation, and slightly lower in intrusiveness.

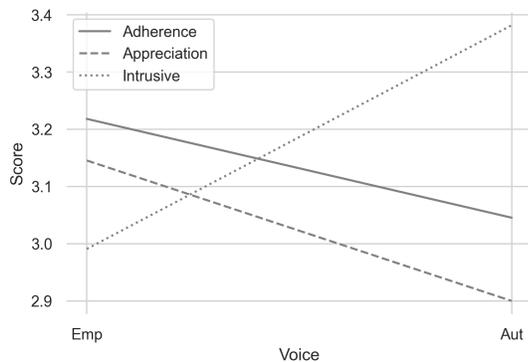


Figure 2: Persuasion scores for the physical activity scenarios

H4: Recommendations that aim to let the users make healthier food-related choices will be more persuasive when the source is empathic. The scores of the two scenario’s that related to food were taken. Then a repeated measures ANOVA was performed to compare the differences between these scores. The results of this test can be found in Table 5.

The results show that there are no significant differences between the empathic and the authoritative voice in terms of adherence, appreciation, and intrusiveness. This means that the hypothesis *H4* is rejected as well. Similar to *H3*, the empathic voice scores only slightly higher in adherence and appreciation, and slightly lower in intrusiveness.

Variable	F-value	p-value
Adherence	0.569	0.452
Appreciation	3.649	0.059
Intrusiveness	2.082	0.152

Table 5: Repeated measures ANOVA for the food scenarios

4.2 Qualitative results

This section showcases the results from the open question at the end of the survey in which participants were asked to share any final thoughts on the research or personal assistant. Of the 111 participants, 33 left an extra comment related to the research or personal assistant. As there were only 33 comments, they were interpreted manually, without usage of any additional tools or coding software.

A representative quote for a positive attitude was: *“I believe that it can be a very valuable tool, especially for elderly people or people with sheltered housing”*. A negative attitude is illustrated by the remark: *“I’m mostly annoyed by what is being said. They are the things you already know and ignore, so an extra voice saying them is very annoying.”* All comments that addressed the context of the recommendation stated that the (emphatic or authoritative) manner in which the recommendation is given impacts the way it is perceived: *“For example, saying that I’m lazy when I’m sitting on the couch is annoying, but if she would say that it is healthy to get some physical activity in, that would be positive.”* One participant stated explicitly: *“Let her express what you should do, not what you should not do. Less the ‘tone’ of your mother, but more that of a ‘wise friend’.”*

4.3 Result summary

For all types of lifestyle-related persuasive recommendations, the empathic voice scored higher in terms of adherence and appreciation, and lower in terms of intrusiveness. Contrary to what is suggested in the literature (see Section 2.4.5), the difference were not significant for food-related recommendations. Furthermore, for recommendations regarding physical activity, the emphatic voice was slightly preferred as well in terms of higher adherence and appreciation, and lower perceived intrusiveness.

This consistent preference for an empathic Voice Assistant, as observed in the quantitative results, is in line with the qualitative results: the various comments all indicated that well-intended persuasive arguments can very easily be perceived as patronizing.

5 INTERPRETATION AND DISCUSSION

In this study, the difference in persuasiveness of a persuasive Voice Assistant with an authoritative voice, versus one with an empathic voice was tested. As the results showed, all of the hypothesis were rejected. Where the hypothesis stated that there would be no significant difference between the two voices, the empathic voice appeared more persuasive – and vice versa.

As became clear in Section 2, there exist quite some contradiction in research related to persuasive technology, in particular on how specific character traits and personalities of the technology affect persuasiveness. It was suggested that this contradiction could very

well be caused by context dependency [25]. This was also supported by participants' comments such as "For me it mostly depended on how it was presented. Especially the first recommendation was a lot more negative than the last one" and "Depending on what the PA responds to and how it does so impacts the chance of acceptance".

The variables in this study were standardized as much as possible, in order to avoid as much noise as possible. Participants were led through a standardized scenario, the possible influence of voice gender was controlled for, and the variations between the empathic and authoritative phrasing were consistent for all six recommendations – in three different contexts. Nevertheless, we could not reproduce the effects as suggested in the literature [27]. Instead, we found that participants preferred the empathic assistant in all situations.

The consistent preference for the empathic Voice Assistant, as observed in this study, may be a context artifact as well: all participants of this study were Dutch and in terms of Hofstede's cultural dimensions [15], the Netherlands score low on power distance, with strong preferences on being independent, 'hierarchy for convenience only' and a high level of individualism⁵.

If context is indeed the cause of the differences in persuasiveness, it is very difficult to come up with a reliable design for a Recommender System, contrary to what is suggested in existing literature [12, 22]. Not only are there many different contexts to take into account for a real world application, what parts of the context are important in the design of a Recommender System is also very unclear.

5.1 Limitations of the study

In this study, there were several aspects that could potentially be improved on or elaborated on more in future research. First of all, the research methods. The decision was made to conduct the research through an online survey in Qualtrics, to standardize the scenarios as much as possible. However, this also offers a few problems, among others that the environment cannot be controlled. As participants took part of the study using their own devices and at a moment of their choice, external factors may have affected the way in which the participants answered the questions. Finally, there is still quite a difference between a simulated and real-world situation and users might respond differently to a Recommender System at different periods during the day.

5.2 Implications

Persuasive recommender systems, either or not implemented as voice assistants, can have a major influence on modern day society. Especially with more and more people having a smart device in their near vicinity, it is very easy to obtain a voice assistant like the one that was described in the scenario's in this study. Such a tool could be very powerful, but also introduces many risks that are important to consider.

First and foremost, as reflected in the comments of our participants, there is a risk that a persuasive voice assistant that is constantly 'on' will be considered annoying or even patronizing.

This would not only diminish or annihilate the effect of the persuasive advices, but most likely would even have a negative effect on general well-being.

Further, following the trend of health insurance companies that increasingly provide their clients with lifestyle apps, as a bonus or in exchange for (financial) profits [31], it is not unlikely that persuasive voice assistants for lifestyle advice will appear soon. As discussed in Section 2.3, this development comes with several ethical challenges that may have an impact on user acceptance.

Building upon the emergence of multi-stakeholder persuasive lifestyle assistants, it is important to keep in mind that commercial products are developed to serve (among others) commercial values, which may or may not be aligned with the interests of the end-user. Following Konstan and Terveen [18], we believe that it is important to keep up with industry and to understand the processes that shape the design of such recommendations and which persuasive techniques are used to influence the daily lives of a potentially large user population.

The results of this study, combined with the observations from existing literature, highlight the limitations of studies that aim to isolate just one factor or just one effect: as the interaction with all other specifics that define the usage context may cancel out any expectations that are based on observations in isolation, it appears to be inevitable to develop more holistic approaches and associated field studies – and to accept that effects may not easily transfer from one context to other contexts.

REFERENCES

- [1] Izak Benbasat and Wei-quan Wang. 2005. Trust in and adoption of online recommendation agents. *Journal of the association for information systems* 6, 3 (2005), 4.
- [2] Alexander Benlian, Johannes Klumpe, and Oliver Hinz. 2020. Mitigating the intrusive effects of smart home assistants by using anthropomorphic design features: A multimethod investigation. *Information Systems Journal* 30, 6 (2020), 1010–1042.
- [3] Gerd Bohner, Markus Ruder, and Hans-Peter Erb. 2002. When expertise backfires: Contrast and assimilation effects in persuasion. *British Journal of Social Psychology* 41, 4 (2002), 495–519.
- [4] Robin Brewer, Casey Pierce, Pooja Upadhyay, and Leeseul Park. 2022. An empirical study of older adult's voice assistant use for health information seeking. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 12, 2 (2022), 1–32.
- [5] Rose G Campbell and Austin S Babrow. 2004. The role of empathy in responses to persuasive risk communication: Overcoming resistance to HIV prevention messages. *Health communication* 16, 2 (2004), 159–182.
- [6] Michelene TH Chi, Robert Glaser, and Marshall J Farr. 2014. *The nature of expertise*. Psychology Press.
- [7] Jason K Clark, Duane T Wegener, Meara M Habashi, and Abigail T Evans. 2012. Source expertise and persuasion: The effects of perceived opposition or support on message scrutiny. *Personality and Social Psychology Bulletin* 38, 1 (2012), 90–100.
- [8] Dominic De Franco, Alison Pease, and Mark Snaith. 2018. Measuring persuasiveness in behaviour change support systems. In *CEUR Workshop Proceedings*, Vol. 2102. CEUR-WS, 41–47.
- [9] Anneke De Graaf, José Sanders, and Hans Hoeken. 2016. Characteristics of narrative interventions and health effects: A review of the content, form, and context of narratives in health-related narrative persuasion research. *Review of Communication Research* 4 (2016), 88–131.
- [10] Dimitri Dojchinovski, Andrej Ilievski, and Marjan Gusev. 2019. Interactive home healthcare system with integrated voice assistant. In *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. IEEE, 284–288.
- [11] Aarthi Easwara Moorthy and Kim-Phuong L Vu. 2015. Privacy concerns for use of voice activated personal assistant in the public space. *International Journal of Human-Computer Interaction* 31, 4 (2015), 307–335.
- [12] Brian J Fogg. 2002. Persuasive technology: using computers to change what we think and do. *Ubiquity* 2002, December (2002), 2.

⁵<https://www.hofstede-insights.com/country/the-netherlands/>

- [13] Murray A Hewgill and Gerald R Miller. 1965. Source credibility and response to fear-arousing communications. (1965).
- [14] Charles K Hofling, Eveline Brotzman, Sarah Dalrymple, Nancy Graves, and Chester M Pierce. 1966. An experimental study in nurse-physician relationships. *The Journal of nervous and mental disease* 143, 2 (1966), 171–180.
- [15] Geert Hofstede. 2011. Dimensionalizing cultures: The Hofstede model in context. *Online readings in psychology and culture* 2, 1 (2011), 2307–0919.
- [16] Karen Jones. 2012. Trustworthiness. *Ethics* 123, 1 (2012), 61–85.
- [17] Sherron B Kenton. 1989. Speaker credibility in persuasive business communication: A model which explains gender differences 1. *The Journal of Business Communication* (1973) 26, 2 (1989), 143–157.
- [18] Joseph Konstan and Loren Terveen. 2021. Human-centered recommender systems: Origins, advances, challenges, and opportunities. *AI Magazine* 42, 3 (2021), 31–42.
- [19] Clifford Ivar Nass and Scott Brave. 2005. *Wired for speech: How voice activates and advances the human-computer relationship*. MIT press Cambridge.
- [20] Harri Oinas-Kukkonen. 2010. Requirements for measuring the success of persuasive technology applications. In *Proceedings of the 7th international conference on methods and techniques in behavioral research*. 1–4.
- [21] Harri Oinas-Kukkonen. 2013. A foundation for the study of behavior change support systems. *Personal and ubiquitous computing* 17, 6 (2013), 1223–1235.
- [22] Harri Oinas-Kukkonen and Marja Harjuma. 2009. Persuasive systems design: Key issues, process model, and system features. *Communications of the association for Information Systems* 24, 1 (2009), 28.
- [23] Daniel J O’keefe. 2015. *Persuasion: Theory and research*. Sage Publications.
- [24] Marta Perez Garcia and Sarita Saffon Lopez. 2019. Exploring the uncanny valley theory in the constructs of a virtual assistant personality. In *Proceedings of SAI Intelligent Systems Conference*. Springer, 1017–1033.
- [25] Lars-Eric Petersen and Jörg Dietz. 2000. Social Discrimination in a Personnel Selection Context: The Effects of an Authority’s Instruction to Discriminate and Followers’ Authoritarianism 1. *Journal of Applied Social Psychology* 30, 1 (2000), 206–220.
- [26] John Rossi and Michael Yudell. 2012. The use of persuasion in public health communication: an ethical critique. *Public Health Ethics* 5, 2 (2012), 192–205.
- [27] Julio J Rotemberg. 2010. *Persuasion and empathy in salesperson-customer interactions*. Technical Report. National Bureau of Economic Research.
- [28] Emre Sezgin, Yungui Huang, Ujjwal Ramtekkar, and Simon Lin. 2020. Readiness for voice assistants to support healthcare delivery during a health crisis and pandemic. *NPJ Digital Medicine* 3, 1 (2020), 1–4.
- [29] Lijiang Shen. 2010. Mitigating psychological reactance: The role of message-induced empathy in persuasion. *Human Communication Research* 36, 3 (2010), 397–422.
- [30] Lijiang Shen. 2011. The effectiveness of empathy-versus fear-arousing antismoking PSAs. *Health communication* 26, 5 (2011), 404–415.
- [31] A Spender, C Bullen, L Altmann-Richer, J Cripps, R Duffy, C Falkous, M Farrell, Tony Horn, James Wiggzell, and Wendy Yeap. 2019. Wearables and the internet of things: Considerations for the life and health insurance industry. *British Actuarial Journal* 24 (2019).
- [32] S Shyam Sundar and Jinyoung Kim. 2019. Machine heuristic: When we trust computers more than humans with our personal information. In *Proceedings of the 2019 CHI Conference on human factors in computing systems*. 1–9.
- [33] Katja Wagner and Hanna Schramm-Klein. 2019. Alexa, are you human? Investigating anthropomorphism of digital voice assistants—a qualitative approach. (2019).
- [34] Jayme L Wilder, Devin Nadar, Nitin Gujral, Benjamin Ortiz, Robert Stevens, Faye Holder-Niles, John Lee, and Jonathan M Gaffin. 2019. Pediatrician attitudes toward digital voice assistant technology use in clinical practice. *Applied clinical informatics* 10, 02 (2019), 286–294.
- [35] Catherine Zambaka, Paula Goolkasian, and Larry Hodges. 2006. Can a virtual cat persuade you? The role of gender and realism in speaker persuasiveness. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. 1153–1162.