People versus technology. Who can make you want to eat healthier?

Elske Borneman Alessandro Ferretti Anika Kok Thomas Pilaet Xander Verstraeten
Department of Industrial Design, Eindhoven University of Technology, the Netherlands
{e.m.borneman, a.ferretti, a.k.kok, t.h.m.pilaet, x.verstraeten}@student.tue.nl

ABSTRACT
Smart home products are on the rise. They are becoming increasingly popular and are more integrated into our lives than ever before. However, research shows that the effects of smart home products are not always as they seem or are intended by their creators. This study dives deeper into the effects of smart home products related to the creation of behavioral awareness, and compares this to behavioral awareness created by humans. The focus is on the interaction between human and computer, which is studied through the parameters credibility, satisfaction and desirability. In order to study the differences between awareness created by an artificial agent and the human, videos visualizing the same scenario were made and used in a questionnaire following the lab methodology. The results of the questionnaire show a notable difference between the human and the artificial agent for all parameters. The human agent is considered to be significantly more credible. It also seems to perform more desirable actions than the artificial agent. Lastly, the human's actions also seems to result in a higher intent for behavior change.

CCS CONCEPTS
•Social and professional topics•General and reference~Document types~Surveys and overviews

KEYWORDS
Generating awareness; Human compared to artificial agent; Smart home; Smart assistant;

INTRODUCTION
As technology devices and artificial intelligence continue to develop, people encounter sophisticated "smart products" in their homes increasingly (Luria, Hoffman, & Zuckerman, 2017). The likelihood of regularly interacting with these "smart products" in people's daily lives is near, if one is not already experiencing it now. However, there is very little knowledge about how one perceives, interacts with or accepts these artificial agents in social contexts (Luria et al., 2017)

Smart home products are marketed to increase convenience and lower energy consumption, which is achieved through a process of simplification and streamlining. However, as Strengers and Nicholls say smart home products do not always live up to what they are claimed to be or help with (Strengers & Nicholls, 2017). At times they even cause the opposite effect of what they are supposed to do. Most notably, they seem to have increased the pace of everyday living, which people now tend to fill with new tasks making them feel even more hurried than before (Strengers & Nicholls, 2017).

In this study, the effects of smart home products are further examined, in order to gain a clearer understanding of how they influence the choices of their users. The insights and results on this can be used to make better choices in the design and advertisement of smart home products. Smart home products are often marketed as life improving products, but little evidence shows this is true. As an intake in researching a possible life bettering possibility from a smart home product, the research looks into how they can influence eating habits through social awareness (Robinson & Field, 2015).

This study aims to evaluate what the differences are between behavioral awareness created by a person and behavioral awareness created by an artificial agent. In order to study the differences between these two types of agents, multiple videos visualizing the same scenario were made and used in a questionnaire following the lab methodology (Abel, 2011). To be able to compare the results properly, the scenario has been kept as similar as possible in both cases, with the only difference being the carrier of the message; in this case, the human waiter versus the artificial waiter.

In this paper, the authors will elaborate on previous research regarding the topic smart home products and the parameters awareness, credibility, satisfaction and intent for behavioral change, which are the main metrics used to evaluate the differences between the awareness created by a person and awareness created by an artificial agent.

RELATED WORK

Smart home products

Since the rapid industrialization, a lot has changed in the average household. Initially this was caused by the emergence of household appliances at around 1920-1930. For the first time, there were washing machines, vacuum cleaners, and other nifty appliances available to the general public (Cowan Schwartz, 2012). Even now, houses are getting more advanced with different data processing objects with interactive functions, named smart home products. All with the promise to improve the home: a place for security and control, for activity, for relationships and continuity, and for identity and values (Gram-Hanssen & Darby, 2018). Besides increasing convenience, these products could also be a cost-effective solution of improving home care for the elderly and the disabled in a non-obtrusive way, allowing greater independence, maintaining good health and preventing social isolation (Chan, Campo, Estève, & Fourniols, 2009).

However, these promises are not necessarily true in reality. Smart home products are marketed to increase convenience and lower energy consumption, but most often don't have the effects as predicted and in some cases even have the complete opposite effect on convenience and fail to save energy. Labor-saving devices have failed over and over to relieve people, especially women, from the time-consuming burden of household chores (Committee & History, 1986). While such devices are intended to save time, smart home products have enabled the increased pace of everyday living, potentially resulting in people feeling more harried and time-pressured. This happens because people tend to fill in the time they saved by introducing a smart home product with new tasks and energy-consuming activities in order to feel like they are using their time efficiently (Strengers & Nicholls, 2017).

Human agents and artificial agents

To define the differences between human and artificial agents, definitions of published studies have been explored. A study by Wachsmuth defines a human agent by its appearance and as someone who is using reasonable natural language dialogue to express his/her desires, goals, and intentions (Wachsmuth, 2008). Human agents also are different by having consciousness, which describes the fact that humans are aware of thoughts and sensations. This is contradicting to artificial agents, who have intentional states and perceive others as intentional agents which is inherently different (Wachsmuth, 2008). A study by Chen Yu gives an overview to understand the difference between the small behavior differences or as he calls them, micro- behaviors, in human-human communication versus robot-human communication (Yu, Schermerhorn, & Scheutz, 2012). In the behavioral type of speech, the results show that humans will use more distinct words, more tokens/words and more one-word utterances. Recent findings from social robotics, virtual reality, psychology, and neuroscience to examine how people recognize and respond to emotions displayed by artificial agents are reviewed by Hortensius (Hortensius, Hekele, & Cross, 2018). This review gives an overview of the current state of emotion expression and perception during interactions with artificial agents and states that humans can, to some extent, accurately perceive the emotions expressed by these artificial agents but this can cause as well positive and negative reactions among the perceivers. While people can feel empathy in case of suffering of the artificial agent, they might also feel aggression towards them in other contexts.

RESEARCH METHOD

This exploratory research is about humans and their relationship and interactions with smart home products. To explore the effects of smart home products, the non-symmetric human-computer collaborations by comparing two interaction styles was researched. This was done by conducting a laboratory (LAB) study in which primarily quantitative data on a five point Likert scale was collected by means of a video-guided questionnaire. This is comparable to previous published studies like for example the study of Stange and Kopp (Stange & Kopp, 2020). This video-guided questionnaire set-up was chosen since it fit the time frame and objectives of the study, namely, validating three hypotheses that were set up prior to the questionnaire going live.

These hypotheses for this study are:

1. Behavioral awareness created by humans has more credibility than an artificial agent.
2. Behavioral awareness raised by a human agent is considered more desirable than when raised by an artificial agent.
3. The intent for behavioral change is higher when behavioral awareness is raised by a human agent.

The objective of this study is to find what the differences are between behavioral awareness generated by a person and behavioral awareness generated by an artificial agent as illustrated in Figure 1. These differences are split up into multiple parameters: intended behavior change, credibility (Appelman & Sundar, 2016; Waddell, 2018) and desirability (Kinch, 2016; Stange & Kopp, 2020). To substantiate the results on desirability, an extra parameter has been added to the questionnaire, namely satisfaction, this is done based on the USE Questionnaire of Arnold Lund (Lund, 2001).

The questionnaire consisted of four sections: an introduction to the questionnaire itself and the general scenario, questions about scenario one, scenario two and a general section where final remarks could be made. The introduction and two scenarios were conceptualized in videos to help participants imagine themselves in the two scenarios used to compare the effects of smart home products. The 2 scenarios were given in a random order to the participants, to overcome possible biases by for example showing everybody always the human first. Without the participants knowing there will be an artificial one after it. At the start of the questionnaire was asked if the participant was born on an even or uneven date, to randomly separate all the participants 50/50. The general scenario shows someone who just went for a run and is now ordering a burger with a side dish in a sports canteen. After the introduction video was shown, the participant was asked what option they themselves would choose as a side dish (salad or fries) with the burger. This introductory is followed up with two different scenarios. One scenario contains a human waiter taking the user’s order. The other scenario contains an artificial waiter taking the order.

The questionnaire contained Likert scale questions which were built up with the nominal characteristics (strongly disagree, disagree, neutral, agree, strongly agree). Each individual parameter was split up in three Likert scale questions. For the analysis of the data the scores for these three questions were added together to get the final score for each parameter. Thus, ranging from 3-15. On top of the Likert scale questions, two open questions needed to be answered by the participant. One of them allowed participants to elaborate on their answers given on the question whether they would change their choice of side dish after having heard the information from the waiter in question.

In the second open question participants could indicate how the waiters could behave differently in the future.

The study aimed for fifty to one hundred participants. The participants were selected based on convenience sampling, with no specific requirements to in- or exclude participants from the research. A short pilot test was conducted with four participants. Afterwards the questionnaire was available for participants over a period of one week. Within this timeframe, 62 questionnaires were completed.

The dataset was checked for missing data and errors before analysis. The data was then analyzed using descriptive statistics and correlations were found through data visualization, in the software Tableau. The normality of the Likert scale scores was tested with a Shapiro-Wilk test. The significance of the differences in the results were tested with either a Wilcoxon signed-rank test or a sign test depending on the symmetry of the data for each parameter.

The results of this study are relevant for researchers and developers within the HRI and HCI sector, because it builds on previous research on the design of social robots and other HRI practices. Furthermore, its results can be relevant for experts designing for the smart home environment as its outcomes could be used for promoting healthy eating behavior in homes.

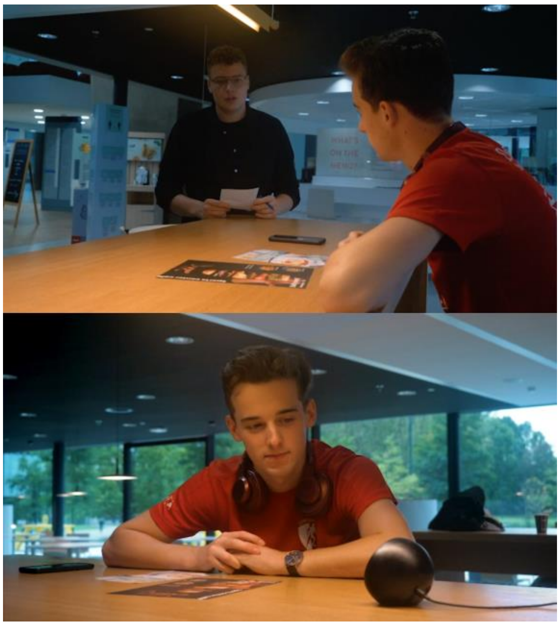

Figure 1. Comparison view of the two waiters in the videos used to visualize the scenario

RESULTS

The exploratory study was conducted in the Netherlands. The goal was to further explore the effects of smart home products on humans through a questionnaire, resulting in quantitative and qualitative data. Within the study 62 participants (36 female, 26 male) were involved, of which 60 were included in the analysis after data cleaning the data. The 60 adults (all 18 years and above) consisted of different age groups where 43 of them were 18-24, 10 of the age of 25-34 and 7 had the age group of 45-64.

Multiple questions as shown in appendix I were implemented to measure the user's experience on the 3 different parameters (credibility, satisfaction, desirability) based on the hypotheses. The scores for these three questions were added together to get the final score for each parameter. As there were three five point Likert ratings for each parameter, the highest score a participant could give to a parameter was fifteen and the lowest three. Table 1 shows the legend for these ratings.

In the visualizations of the Likert scale ratings, percentages are also shown within each colored area. These percentages show what amount of the total participants rated the waiter that score for the parameter in question.

	Combined rating	Total Likert score
	Very high	14-15
	High	11-13
	Moderate	8-10
	Low	5-7
	Very low	3-4

Table 1. Legend of total Likert ratings of parameters

Credibility

Table 1 shows the mean and standard deviation of the questions A, B and C that are linked to the aspect of credibility. Participants rated the human waiter's credibility an average of 3,894 where they rated the artificial agent a 3,750. Figure 3 and 4 show the difference in ratings for credibility for the two types of waiters.

	Human	Agent	Human	Agent
Question	*Mean*		*SD*	
A	3,817	3,800	0,866	0,909
B	3,783	3,467	0,732	0,903
C	4,083	3,983	0,526	0,741

Table 2. Credibility result statistics

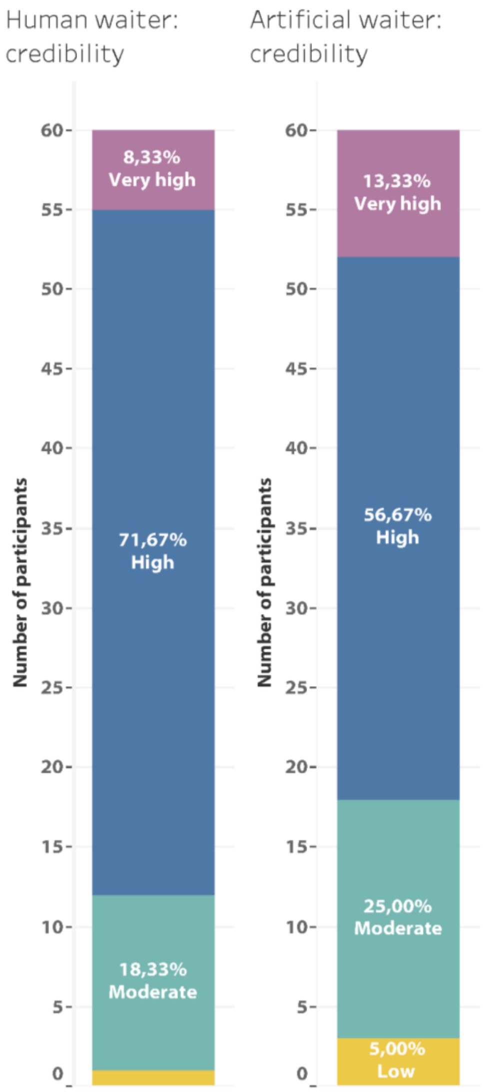

Figure 2. Credibility ratings for the waiters

Satisfaction

Table 2 shows the mean and standard deviation of the questions D, E and F that are linked to the aspect of satisfaction. Participants rated the human waiter's satisfaction an average of 2,794 where they rated the artificial agent a 2,628. Figure 5 and 6 show the difference in ratings for satisfaction for the two types of waiters.

	Human	Agent	Human	Agent
Question	*Mean*		*SD*	
D	2,800	2,467	0,909	1,024
E	2,567	2,733	1,086	0,873
F	3,107	2,683	0,866	0,975

Table 3. Satisfaction result statistics

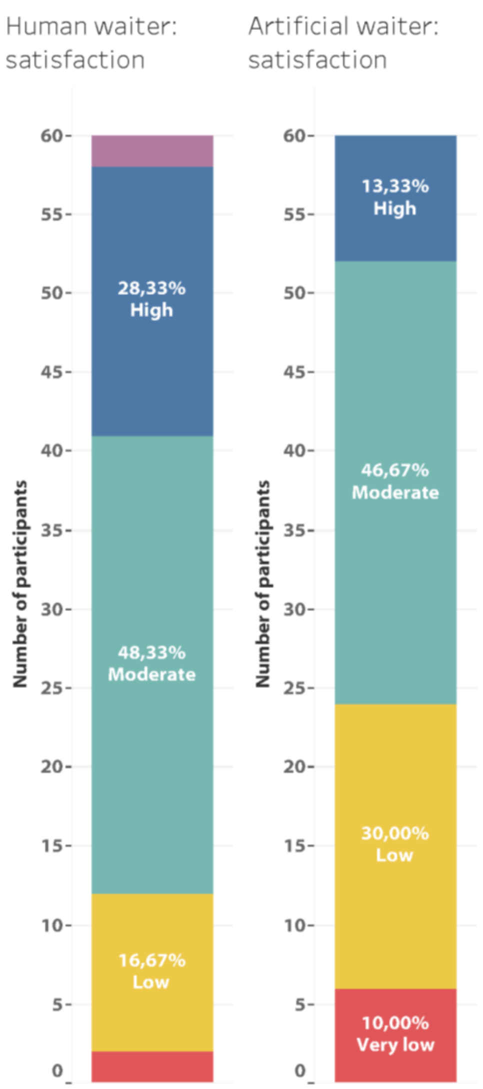

Figure 3. Satisfaction ratings for the waiters

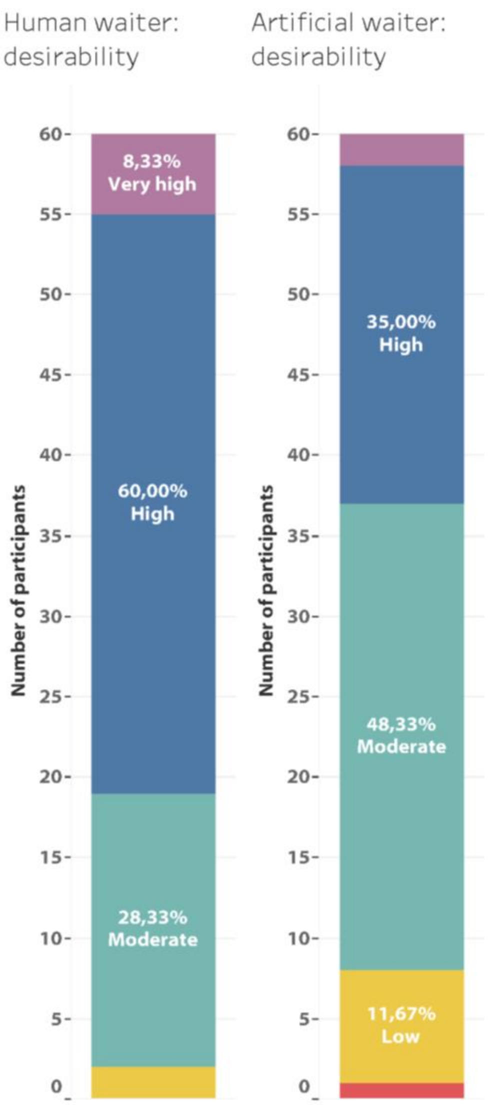

Figure 4. Desirability ratings for the waiters

Desirability

Table 3 shows the mean and standard deviation of the questions G, H and I that are linked to the aspect of desirability. Participants rated the human waiter's desirability an average of 3,428 where they rated the artificial agent a 3,161. Figure 7 and 8 show the difference in ratings for desirability for the two types of waiters.

	Human	Agent	Human	Agent
Question	*Mean*		*SD*	
G	3,567	3,367	0,803	0,912
H	3,233	2,783	0,901	1,002
I	3,483	3,333	0,764	0,925

Table 4. Desirability result statistics

Intent for behavior change

Looking at the mean values for the responses on the question whether the participants would change their choice of side dish after hearing the information from the waiter, the mean value for the human waiter was 3,27 and for the artificial waiter 2,98. Figure 9 and 10 show the difference in ratings for intent for behavior change for the two types of waiters. Figure 8 shows the legend for the different colors used in the visualizations of the results.

Legend	
	Strongly agree
	Agree
	Neutral
	Disagree
	Strongly disagree

Table 5. Legend of intent for behavior change ratings

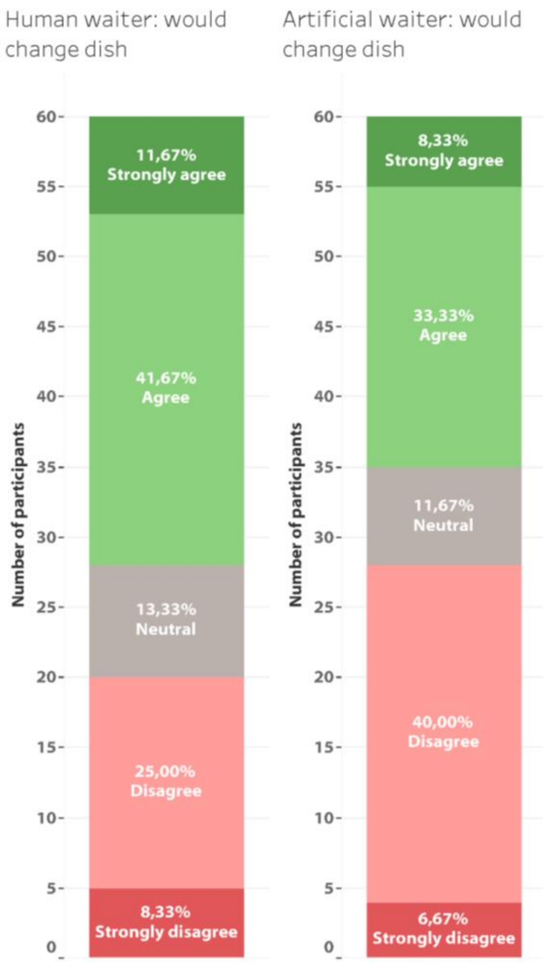

Figure 5. Ratings for changing choice of side dish after interactions with the waiters

When asked to elaborate on their choice, given in the above question. Multiple participants get the feeling that the human waiter has more of a genuine approach than the robot waiter, that felt more pre-programmed. One participant (male, 25-34) elaborated on how the social impact of a real person might influence him more:
"Talking to a real person adds pressure. Especially if I already was slightly overweight I would not dare to pick the perhaps more tasty fries (this is debatable)". The same goes for not seeing the robot as a real entity with consequences towards people's own appearance on others: *"It is easier to say no to a robot as it can't make you feel as bad"* (female, 18-24).
Asking the participants if they had a smart speaker (artificial agent) and how confident they felt when talking to it resulted in nominal five point Likert scale data. 48 out of 60 participants did not have a smart speaker. Where 22 people did felt a high confidence and 23 people a moderate level of confidence when talking to a smart speaker.

Generalization of the data to the population

The normality of the Likert scale scores for all parameters was looked at with a Shapiro-Wilk test. This test showed that most scores were significant. Therefore, it was not possible to assume normality for all scores. With that, the assumption of normality was not met to conduct a paired t-test.
For this reason it was decided to conduct a Wilcoxon signed-rank test for all Likert scale scores. To do this the distribution of the differences between the scores for the human waiter and the scores for the artificial waiter needed to be symmetrical in shape. This was tested through the analysis of boxplots of the data from the different parameters. It was found that the data for credibility and desirability was symmetrical. The data for satisfaction and intent for behavior change was not. Therefore, a Wilcoxon test was conducted for credibility and desirability. A sign test was conducted for satisfaction and intent for behavior change.

A Wilcoxon signed-rank test revealed that the Likert scale scores for credibility for the human waiter did show a statistically significant higher score compared to the scores for credibility for the artificial waiter ($Z = -2{,}114$, $p = 0.035$).
A sign test revealed that the Likert scale scores on satisfaction for the human waiter did not show a statistically significant higher score compared to the scores for satisfaction for the artificial waiter ($Z = -1.556$, $p = 0.120$).
A Wilcoxon signed-rank test revealed that the Likert scale scores on desirability for the human waiter did show a statistically significant higher score compared to the scores for desirability for the artificial waiter ($Z = -2{,}565$, $p = 0.010$).
A sign test with binomial distribution revealed that the Likert scale scores on intent for behavior change for the human waiter did not show a statistically significant higher score compared to the scores for the intent for behavior change for the artificial waiter ($p = 0.052$).

CONCLUSION

This research aimed to evaluate the differences between behavioral awareness generated by a person and behavioral awareness generated by an artificial agent. This was done in order to gain a better understanding of the effects of smart home products (on its users). Two scenarios were designed in which a person would order food in a sports canteen after going for a run. In one of the two scenario's their waiter was human and in the other artificial. The scenarios were shown as videos in a questionnaire after which the participants were asked to rate the shown interactions between the athlete and the waiters. The participants rated the interactions on credibility, satisfaction, desirability and intent for change of behavior.

For hypothesis 1, Figure 2 and the Wilcoxon test on credibility indicate that the human waiter was considered significantly more credible than the artificial waiter. Therefore it can be concluded that for the scenario in the study behavioral awareness created by humans has more credibility than an artificial agent, which means that hypothesis 1 can be confirmed. Interestingly, as Figure 2 indicates, both waiters were considered to be highly credible, with only a small margin being moderate or low rates.
For hypothesis 2, the sign test on satisfaction did not show a significant difference between the two types of waiters. The Wilcoxon test on desirability and also Figure 4 do show a significant difference between the desirability ratings of the waiters. As satisfaction should have been an extra validation for the ratings of desirability, hypothesis 2 can only be partly confirmed.
Lastly, for hypothesis 3 the sign test on intent for behavior change did not show a significant difference between waiters. While Figure 5 does show that participants were more inclined to change their behavior for the human waiter, hypothesis three cannot be confirmed with certainty. Leaving it also only partly confirmed.

In summary, the conclusion can be made that there is indeed a difference between behavioral awareness generated by a person and behavioral awareness generated by an artificial agent. A human agent significantly performs better for credibility. There is also an indication that a human agent's actions seem to be more desirable. Lastly, a trend can also be seen that a human agent seems to generate more intent for behavior change.

DISCUSSION

In this section the results and the interpretations of the researchers will be discussed. The hypotheses and the limitations of the results will be elaborated on.

The confirmation of the three hypotheses implies that smart home products such as the voice assistant in this case, could influence their user's behavior, because they think the information given is not only credible but also desirable. Our research shows that higher ratings on credibility, desirability and satisfaction lead to a higher intent for behavior change, which was clearest for the artificial waiter. In this case, the intent for behavior change related to food consumption was measured but the influence that artificial agents have on our behavior could extend beyond that.

It appears that people are getting more and more used to having artificial agents around which could result in more (unintended) side effects than can be foreseen right now (Strengers & Nicholls, 2017). From this study can be concluded that there is only a small difference in the credibility of humans and artificial agents. This is confirmed in other research by Fan as studies show that participants can consistently tell the difference between artificial agents with a high and low emotional intelligence, and that agents with high emotional intelligence are rated as more trustworthy (Fan, Scheutz, Lohani, McCoy, & Stokes, 2017). This raises new questions such as: what would happen if the artificial agent becomes more credible or trustworthy than the human? Or: What if the human is perceived more credible, while the artificial agent is communicating correct information and the human is giving incorrect information? To what extent are people already being influenced and how will this progress in the future? All questions that need to be further studied in order to gain a better understanding of the effects of the ever changing presence of artificial agents in our lives.

In any case, both waiters scored very moderately on satisfaction. Their desirability ratings were also considerably lower compared to their credibility ratings. These lower ratings could be connected to the unexpected behavior of the waiters when naming the calories in the dish dishes. Something that would not happen normally in a restaurant. This does leave room to wonder what would happen in case information or advice is given at a more logical, wanted moment. It is expected that the desirability rating of receiving information from an artificial agent rises if a person is looking for certain information.
The possible reason as to why the human waiter is favored over the artificial waiter, can come from a source of social cohesion with the waiter. Multiple participants stated that they felt more connected to the human compared to the artificial waiter. This connection will likely be a collection of different feelings that occur with a human and not with a robot yet. Giving the human more influence on the people for their side dish choice. For example: the judgement of another human can have an influence on the behavior of the person receiving information. In the current situation, these feelings of judgement appear to be much less apparent in the interaction with the artificial agent, which most likely influences the intent for behavior change.

Limitations

During the set-up of this lab research, a few assumptions were made. This was done to work within the limited time that was available within this research project. In this chapter, it will be explained which assumptions were made, and elaborate on how these might have affected the results of the study.

Most importantly when the questionnaire was published it was assumed that the three terms used for each of the three parameters (credibility, satisfaction, desirability) would be reliable. This assumption was made as the terms for each parameter were grounded in literature. The reliability of the terms was tested only after the data was analyzed. This was done with a Cronbach's alpha test. This test showed that the three terms for credibility seemed only to be reliable for the artificial waiter. However, as these terms were already strongly based in literature it can still be assumed that the terms used for credibility were reliable [21, 22]. The test indicated that the terms used for satisfaction were reliable. Lastly, for desirability it indicated that the used terms were not reliable. The found reliability could be caused by the small sample size and might be mitigated by increasing it.

One of the other assumptions is that the choice for only a male waiter and smart assistant voice would not affect the outcome of the research in a crucial way. This is why both male and female participants saw the same videos. It was assumed that female participants can just as well identify themselves with the athlete shown in the videos as the male participants can. Another assumption made is that the participants will have a basic understanding of smart products and are not unfamiliar with seeing a human talk with a smart device. This was checked in the demographics section of the questionnaire. Lastly, most participants probably do not have English as their first language, due to the convenience sampling recruitment method used in this study. However, the assumption was made that the English proficiency level was high enough to understand the definitions used in the survey.

The waiter in our research was computerized to minimize the difference between the very robotlike artificial waiter and the possibly extravagant or emotional human waiter. In this research, both waiters were talking about the exact amount of kcal in the side dishes. Normally, it would be very rare to hear someone talking in this way. There is also a lack of personal character or emotions, which are things that differentiate humans from computers. This is done in order to make scenarios more parallel for comparison, but also limits the study in the accuracy of the outcome compared to reality. A paper written by already indicated that a teacher as robot and a robot as teacher could both be rated as credible (Edwards, Edwards, Spence, Harris, & Gambino, 2016). However, they also indicated that students still found a teacher as a robot more credible and affectionate, despite controlled instructional performances. The results from our own study, combined with the findings of Edwards could imply that artificial agents can be credible and desirable in some cases, even though they are computerized and without personality. This could happen because in some cases, getting straight facts without emotion is wanted.

Another factor the authors were unable to take into account was the mood and the preconception of the participant towards the human waiter in question (the actor). In case the participant was not convinced by the expert level of the waiter, the voice assistant might have felt like a more credible option in this case. This could however also work vice versa; if the waiter was an expert visibly or noticeably, this could have positively influenced the credibility of the human waiter in relation to advice on healthy food consumption. Also, if there were any relevant external influences, such as the mood of the participant, this could have had a big impact on the responses in the questionnaire. For example: if the participant felt really hungry, or if the participant was on a healthy diet anyway. Especially in the latter case, the choice for the salad would have been more obvious, even without the information given by the waiters. A qualitative study that includes more extensive argumentation for the choices made by the participants could be a next step in validating the results from this quantitative research.

When setting up the research, the familiarity with the waiters has been taken into account. In most cases, a human-computer relationship is neutral, meaning there is little bias towards the information given by the artificial agent. However, human-human relationships include much more elements that can influence the opinions of people. For example: a friend that you know is unhealthy might be perceived as less credible than the fitness coach at the sports center. In order to take away as much of these influencing factors, the study was taken out of the smart home context and brought into the public (restaurant) space, where it would make more sense to get information on food consumption by someone unknown. Further research back in the smart home is needed in order to make relevant and valid statements on how artificial agents can influence our behavior in that context. This research could for example incorporate the intelligent agent into the household, where it needs to take a different role than a waiter. People could be remembered about nutritional values when they for example ask the artificial agent for different kind recipes, or with food options that are not familiar with the user. A follow-up study can be executed where the consumption of food is taken out of the equation for behavioral awareness of people when interacting with an artificial agent. The artificial agent could make remarks on house chores within the home setting. Where afterwards it can be measured if the artificial agent increased the amount of time spent on house chores.

REFERENCES

Abel, C. (2011). Design research. In *CTBUH Journal*.

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

Chan, M., Campo, E., Estève, D., & Fourniols, J. Y. (2009). Smart homes - Current features and future perspectives. *Maturitas*, *64*(2), 90–97. https://doi.org/10.1016/j.maturitas.2009.07.014

Committee, C., & History, L. (1986). *Review Reviewed Work (s): More Work for Mother : The Ironies of Household Technology from the Open Hearth to the Microwave by Ruth Schwartz Cowan Review by : Linda Kealey Published by : Canadian Committee on Labour History and Athabasca University Press Stable URL : https://www.jstor.org/stable/25142643*. *17*, 358–360.

Cowan Schwartz, R. (2012). The “industrial revolution” in the home: Household technology and social change in the 20th century. *Domestic Ideology and Domestic Work*, *17*(1), 375–397. https://doi.org/10.1515/9783110968842.375

Edwards, A., Edwards, C., Spence, P. R., Harris, C., & Gambino, A. (2016). Robots in the classroom: Differences in students’ perceptions of credibility and learning between “teacher as robot” and “robot as teacher.” *Computers in Human Behavior*, *65*, 627–634. https://doi.org/10.1016/j.chb.2016.06.005

Fan, L., Scheutz, M., Lohani, M., McCoy, M., & Stokes, C. (2017). Do we need emotionally intelligent artificial agents? First results of human perceptions of emotional intelligence in humans compared to robots. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, pp. 129–141. https://doi.org/10.1007/978-3-319-67401-8_15

Gram-Hanssen, K., & Darby, S. J. (2018). “Home is where the smart is”? Evaluating smart home research and approaches against the concept of home. *Energy Research and Social Science*, *37*(October 2017), 94–101. https://doi.org/10.1016/j.erss.2017.09.037

Hortensius, R., Hekele, F., & Cross, E. S. (2018). The Perception of Emotion in Artificial Agents. *IEEE Transactions on Cognitive and Developmental Systems*, *10*(4), 852–864. https://doi.org/10.1109/TCDS.2018.2826921

Kinch, N. (2016). Value, Meaning and Engagement: The 3 pillars of great experience. Retrieved October 24, 2020, from Medium website: https://medium.com/@NathanKinch/value-meaning-and-engagement-the-3-pillars-of-great-experience-51ac6fa53570

Lund, A. M. (2001). Measuring usability with the USE questionnaire. *Usability Interface*, *8*(2), 3–6.

Luria, M., Hoffman, G., & Zuckerman, O. (2017). Comparing social robot, screen and voice interfaces for smart-home control. *Conference on Human Factors in Computing Systems - Proceedings*, *2017-May*, 580–592. https://doi.org/10.1145/3025453.3025786

Robinson, E., & Field, M. (2015). Awareness of social influence on food intake. An analysis of two experimental studies. *Appetite*, *85*, 165–170. https://doi.org/10.1016/j.appet.2014.11.019

Stange, S., & Kopp, S. (2020). Effects of a social robot’s self-explanations on how humans understand and evaluate its behavior. *ACM/IEEE International Conference on Human-Robot Interaction*, (Section 2), 619–627. https://doi.org/10.1145/3319502.3374802

Strengers, Y., & Nicholls, L. (2017). Convenience and energy consumption in the smart home of the future: Industry visions from Australia and beyond. *Energy Research and Social Science*, *32*, 86–93. https://doi.org/10.1016/j.erss.2017.02.008

Wachsmuth, I. (2008). “I, Max” - Communicating with an artificial agent. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *4930 LNAI*, 279–295. https://doi.org/10.1007/978-3-540-79037-2_15

Waddell, T. F. (2018). What does the crowd think? How online comments and popularity metrics affect news credibility and issue importance. *New Media and Society*, *20*(8), 3068–3083. https://doi.org/10.1177/1461444817742905

Yu, C., Schermerhorn, P., & Scheutz, M. (2012). Adaptive eye gaze patterns in interactions with human and artificial agents. *ACM Transactions on Interactive Intelligent Systems*, *1*(2). https://doi.org/10.1145/2070719.2070726

APPENDIX

Appendix I - Questionnaire questions that were asked for each waiter

Likert scale questions

A: "This waiter gives accurate information."
B: "The information given by the waiter was authentic (meaning: genuine or real)."
C: "The waiter was believable."
D: "It is fun to interact with the waiter"
E: "The waiter performs in the way I want him to perform."
F: "The interaction with the waiter is pleasant."
G: "The interaction with the waiter had value."
H: "It was engaging to listen to the waiter."
I: The interaction with the waiter was meaningful."
J: "The information given by the waiter would change my choice of side dish"

Open questions

Explain your answer to statement J.
Please give instructions on how this waiter could behave differently in the future.

Appendix II - Links the videos used in the questionnaire

Intro scenario: https://www.youtube.com/watch?v=jjhjd4D5nOY
Human waiter: https://www.youtube.com/watch?v=m_mDBz18zYE
Artificial waiter: https://www.youtube.com/watch?v=NDHSVLBqYVE

Appendix III – Data validation

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Human change dish	,272	60	,000	,874	60	,000
Robot dish	,264	60	,000	,861	60	,000
Human credibility full	,209	60	,000	,918	60	,001
Robot credibility full	,176	60	,000	,939	60	,005
Human satisfaction full	,117	60	,041	,978	60	,354
Robot satisfaction full	,143	60	,004	,960	60	,049
Human desirability full	,126	60	,020	,944	60	,008
Robot desirability full	,166	60	,000	,961	60	,056

a. Lilliefors Significance Correction

Reliability test – Cronbach's Alpha (higher than 0.7 is reliable)

Human waiter - credibility

Cronbach's Alpha	N of Items
,626	3

Artificial waiter - credibility

Cronbach's Alpha	N of Items
,733	3

Human waiter - satisfaction

Cronbach's Alpha	N of Items
,792	3

Artificial waiter - satisfaction

Cronbach's Alpha	N of Items
,794	3

Human waiter - desirability

Cronbach's Alpha	N of Items
,681	3

Artificial waiter - desirability

Cronbach's Alpha	N of Items
,656	3

Wilcoxon test (credibility & desirability)

Test Statistics[a]

	Robot credibility full - Human credibility full	Robot satisfaction full - Human satisfaction full	Robot desirability full - Human desirability full	Robot dish - Human change dish
Z	-2,114[b]	-1,369[b]	-2,565[b]	-2,575[b]
Asymp. Sig. (2-tailed)	,035	,171	,010	,010

a. Wilcoxon Signed Ranks Test

b. Based on positive ranks.

Sign test (satisfaction & behavior change)

Test Statistics[a]

	Robot credibility full - Human credibility full	Robot satisfaction full - Human satisfaction full	Robot desirability full - Human desirability full	Robot dish - Human change dish
Z	-2,109	-1,556	-1,680	
Asymp. Sig. (2-tailed)	,035	,120	,093	
Exact Sig. (2-tailed)				,052[b]

a. Sign Test

b. Binomial distribution used.

Appendix IV - Background and contribution of the different authors

<u>Elske Borneman</u>
I graduated from my bachelors, Industrial Design Engineering, from the Applied University of Saxion in Enschede in 2019. After which I started the pre-master of ID last February and continued to the masters this quartile. My most important contributions to this project have been:

- Overall participation to the project in meetings and brainstorms and taking minutes or notes.
- Writing parts of the draft paper with the focus on the related works, the discussion limitations and conclusion.

<u>Alessandro Ferretti</u>
I did my bachelor's degree in Industrial Product Design at the University of Ferrara, Italy. Currently I am M1.1 and this is one of my first courses of my master program. During my bachelor, I worked in a more practical way with little research, but I never wrote a report. My knowledge in this field is very low, therefore I am here to develop my skills and gain a broader experience. My contributions to this project have been:

- Overall participation to the research to meetings and brainstorming
- Writing parts of the paper focusing on abstract and introduction

<u>Anika Kok</u>
I previously studied HBO-ICT & Media Design at Fontys University of Applied Sciences. I did the pre-master at Industrial design and am currently a M1.2. student. My experience is mostly on the practical side of digital design, whereas I am following the master to improve my skills as a designer, but mostly as a researcher and writer. My most important contributions to this project have been:

- Taking a lead role in the first phase of the project, where we defined the research outline;
- Making the videos for the questionnaire;
- Writing the paper and reviewing everything critically with the knowledge that I gained during the (pre-) master.
- Overall participation to the research to meetings and brainstorming

<u>Thomas Pilaet</u>
At the end of last year I received my bachelor degree in Industrial Design at the TU/e. During the bachelor I learned that I want to focus my development on Human-AI collaboration and entrepreneurship. Therefore I will focus myself on MDC, US and BE and follow the DLE track during my masters. I chose the topic of smart homes as it applied to my current M1.1 project in the DIGSIM squad and also to my interests in HCI. At the start of the course I noticed that I seemed to be the most experienced in research and academic writing (together with Anika).

- Advisory role to Anika in the beginning of the course
- Leading role during set-up of questionnaire and analysis
- The gist of the data analysis and validation
- Paper writing
- Overall participation to the research to meetings and brainstorming

<u>Xander Verstraeten</u>
I graduated from my bachelors of Engineering at the HZ University of Applied Sciences in 2019. After which I started the pre-master of ID last February and continued to the masters this quartile. My strengths are in realizing physical prototypes but got a lot of my research skills needed for this course within my pre-master My most important contributions to this project have been:

- Brainstorming towards research possibilities.
- Analyzing the data from the questionnaire.
- Writing the report.