Conformity with Conversational Assistants

Donna Schreuter
Graduation Thesis
Media Technology MSc program, Leiden University
Primary advisor: Peter van der Putten
Secondary advisor: Maarten H. Lamers
d.e.schreuter@umail.leidenuniv.nl

***Abstract*—AI voice assistants are becoming increasingly popular in consumers' daily lives. They are designed and developed to be more humanlike in their speech. This study investigates the effect this can have on ones conformity with an AI assistant. Solomon Asch's series of psychological experiments in the 1950s demonstrated the power of conformity in social groups, and were later replicated with embodied robots. In light of our increasing reliance on AI assistants, this study investigates to what extent an individual will conform to a disembodied assistant. Furthermore, it will investigate if there is a difference between a group that interacted with an assistant that communicates through text, one that has a robotic voice and one that has a humanlike voice. Participants completed a general knowledge quiz with the help of this AI assistant. The assistant would attempt to subtly influence the individuals final responses. We measure how often participants changed their answer to conform with the assistant. Results show a significant difference in conformity between the three groups. Participants conformed significantly more often to the assistant with a human voice than the one that communicated through text. While there is no significant difference between the two voice assistants, the results show that people are more likely to conform with an artificially intelligent assistant when it has a human voice.**

***Index Terms*—conversational agents, social robots, voice assistants, conformity.**

I. Introduction

Artificially intelligent voice assistants, such as Siri, Amazon Echo and Google Home, are designed and developed to be more humanlike in their speech, which allows for a more seamless interaction. This, in turn, allowed these systems to become increasingly popular in consumers' daily lives. The way these systems speak affects the interaction in multiple ways. While some researchers think making AI sound more human has a positive effect on human-robot interaction, others highlight downsides of this. For example, if an AI starts to show speech disfluencies and uses conversation fillers to convince people it is humanlike, people might expect human-like intelligence from it. An iconic example is Google Duplex: a virtual assistant that sounds so human it can complete real-world tasks over the phone, such as making restaurant reservations and hair salon appointments [1].

If people were to act more casual around the AI and start to show speech disfluencies, the AI might not be able to process this. There is also the risk of an uncanny valley of speech [2]. Instead of trust and intimacy, a human-like speaking AI might evoke eeriness. This study investigates the effect this can have on one's conformity with an AI assistant and discusses whether developing the assistants to sound more human pays off.

In the 1950s Solomon Asch performed a series of psychological experiments on conformity. He demonstrated the power of conformity in groups. Asch looked at how pressure from a social group could lead people to conform, even when they knew that the rest of the group was wrong. His experiments showed that 75% of participants conformed to the group at least once and would give an incorrect answer, even when they knew the correct answer.

Previous studies have replicated Asch's conformity experiments with robots [3][4][5], but in most cases these robots were embodied and physically in the same room as the test subject. These experiments tested if robots could have the same social conformity effect as a group of people. However, we believe that current consumer-driven developments of conversational assistants ask for a focus on the effect of voice in these interactions.

In general, people are more likely to interact with a virtual assistant in their daily lives than with an embodied robot. As most smartphones have some kind of artificial intelligence system people can interact with, there is the subsequent rise in voice assistants present in modern households.

No research has been done yet on the conformity effect these single disembodied assistants can have on people. In the light of increasing popularity of virtual voice assistants in consumers' daily lives, a similar study on conformity with these voice assistants is relevant.

This question posed in this study is to what extent people conform to a human-like or robotic sounding AI voice assistant when answering a series of multiple-choice questions under time pressure. We hypothesize that people will conform with the assistant in all conditions. We expect a difference between virtual assistants with a voice and the virtual assistant that does not have a voice.

We conducted an experiment where participants completed a general knowledge quiz with the help of a virtual assistant that communicated through either text, a robotic voice or a humanlike voice. The assistant would attempt to subtly influence the individual's final responses. We measured how often participants changed their answer to conform with the assistant under all three conditions.

This experiment seeks to illustrate what conditions, if any, lead people to conform to an AI. Particularly, whether the presence of a voice leads to more conformity and if there is a difference between a computer-generated robotic voice and a human voice. The aim of this study is not to determine whether

conformity with a conversational assistant is a positive or negative development: this depends on the field from which it is studied and the values that are inherent to it.

Finally we will discuss how the results relate to earlier demonstrated conformity to a group of other people.

This paper is structured as follows: in section II an overview of the background and related work is given. This will discuss the theory of conformity, acceptability of robots, earlier studies that combine the two and finally, the influence the presence and sound of a voice can have on human robot interaction. In section III, the method is explained. It describes the experiment and data-analysis. The results of the experiment and the statistical analysis are stated in section IV, followed by the discussion of the results in section V. The discussion reflects on the results in relation to the background and related work. Finally, answers to the research questions are discussed in the conclusion, along with suggestions for future research.

II. Background and Related Work

This section will first give an introduction on voice-enabled conversational agents and factors that influence these agents' acceptance by humans. It explains why a certain level of trust is important, but also what possible (negative) consequences should be taken into account along this path to acceptability. If people believe machines are better at some tasks than humans, this sometimes results in people letting those machines influence them or even make decisions for them. The importance of transparency and expectation management will be described. This study investigates the relation between speech in virtual assistants and willingness to conform. The theory of conformity by Solomon Asch is used as a starting point to describe social pressure of groups on the individual [6]. This theory is used by other researchers to study similar patterns in Human Machine Interaction. Finally, we define relevant concepts about the voice and the conversational agents' style of speaking.

A. Acceptance of Social Robots

Social robots are designed to interact with humans, often in a humanlike way. Their success relies both on their ability to fulfill certain tasks, and on acceptance from humans who work with them. An obvious way to achieve acceptance is by simulating human appearance and behavior. However, according to Masahiro Mori, the "Uncanny Valley" poses a challenge for the path to acceptability [2][7]. The uncanny valley theory describes the relationship between the degree of human likeness of an object and a person's response to it. Mori hypothesized that a person's response to a humanlike robot would abruptly shift from empathy to revulsion as the robot becomes more humanlike. This descent in the human observer's affinity is called the uncanny valley [2].

Researchers disagree on what triggers the uncanny valley-effect in robots. Some argue that it is caused by appearances, supporting this with studies that show uncanny valley in cartoon images and robots that are not androids. It is also shown that a mismatch in the human realism of a character's face and its voice causes it to be evaluated as eerie [8]. One way in which human and robot differ strongly is the physiology and function that makes the production of speech possible. Speech is, in most speaking robots, generated by TTS (text-to-speech) systems and does not simulate human speech perfectly.

Other factors could trigger the uncanny valley. For example, Gray and Wegner suggest that humanlike robots are unnerving not so much because of their humanlike appearance but more because their appearance prompts "perception of mind", a feeling that the robot can feel and experience things [9]. This is an example of research involving anthropomorphism: the attribution of human traits, emotions, or intentions to non-human entities. Feelings of perception of mind are linked to feeling uneasy. Machines that have experience and a mind of their own are a popular subject in (often dystopian) science fiction, perhaps because it makes people feel uneasy.

Another important factor in acceptance of social robots is trust. While it is important that there is a natural interaction between a human and a social robot, it is even more important for the person to understand the limits of the robots capabilities. As we put more trust in information provided by technology, we get more vulnerable to integrity risks. When a robot shows more human-like features, acceptance might increase. This allows for people to trust it to make decisions or let it influence them on a variety of levels and areas. Previous studies have researched whether people are willing to conform to robots and other non-human agents [4][5][10]. An overview of the most important results will follow. First, it is essential to understand the theory of conformity as this is observed in humans.

B. Conformity with Social Groups

Conformity is the act of matching attitudes, beliefs, and behaviors to group norms or politics. Kelman proposed a social influence theory in which he distinguished three types of social influence or conformity: compliance, identification, and internalization [11]. He defined this social influence as the process in which an individual's attitudes, beliefs, and subsequent actions or behaviors are influenced by referent others, through the three mentioned processes of influence (or conformity types) [11]. While these three types are not mutually exclusive, it is important to understand the difference and more importantly, what drives a change in behavior and attitude in different types of conformity:

- *Compliance* occurs when an individual accepts influence from another person or group to gain approval or avoid being rejected, while possibly keeping their own original beliefs for themselves [11].
- *Identification* occurs when an individual accepts influence from someone who is liked and respected. This type of conformity is motivated by attractiveness of the source. The individual wants to establish or maintain a satisfying self-defining relationship with another person or group [11].
- *Internalization* occurs when an individual accepts the beliefs and behavior of another person or group, and conforms with it, if the source is credible. The individual not

only changes their behavior to fit in with other people, they also agree with them privately or internally [11].

The differences between these three types of conformity can help us understand the scope of the concept. It important to understand that conformity can occur on different levels. In some cases, people genuinely believe another group or individual to the extent that they change their own mind. This is the most concerning type of conformity, since people, groups or perhaps robots can effectively influence other people. In other cases, people behave in accordance with a social group without changing their own internal beliefs. This type of conformity is central to one of the most foundational studies on conformity, conducted by Solomon Asch in the 1950s. Asch generated a disagreement between an individual and a group [6]. He conducted a series of experiments in which participants were asked to complete a number of simple tasks in the presence of a group of 7 informed confederates who were instructed to answer in a predefined pattern. The tasks consisted of 18 comparisons: the participants were instructed to match the length of a line with one of three other lines. One of the three lines was equal to it, the other two were different [6].

Participants were gathered in a room together with the group and gave their answers publicly. They were second to last to answer, allowing the six people before them to create social pressure by unanimously choosing the incorrect answer. 37% of the time, participants would answer incorrectly if the rest of the group did, even when they knew the answer was incorrect. 75% of participants conformed at least once.

The results showed that the answer of a unanimous majority affected the decision-making of individuals. Even though they knew the answer was incorrect, they chose to conform with the group opinion, which indicates that human decision-making can be significantly biased by the presence of a social group that consistently agrees on a certain type of answer [3]. Asch's research focuses on the effect of a group of people, but there are no reasons to believe that people actually believed the answer of the group internally. They might have doubted their own answer at some point during the test, but it is also probable they did not want to attract attention to themselves by standing out of the group. This indicates that they were motivated by a desire to gain approval and a fear of being rejected by the group [11]. Asch found that conformity increased as the group size increased. However, when the group size reaches 4-5, there is little change in conformity [12]. While there seems to be a difference between a group of 2 people and a group of 4, he did not experiment with the effect of a single confederate.

This study investigates whether the same effect can occur when there is just one (non-human) agent creating social pressure. As stated above, different types of conformity with corresponding motivations could occur. Conformity with conversational assistant does not have to be of the same type as conformity with social groups as in Asch's experiments.

C. Conformity with Non-Human Agents

Previous studies have replicated Asch's experiment with a group of robots, but most of them have failed to observe conformity. However, there are studies that show people conform with robots. These studies are variations on Asch's conformity experiments and focus on specific aspects of the experiment or the robots that might affect conformity. For example, the degree of humanness or the type of questions.

Prior studies did not show conformity with robots when there was an objective correct or incorrect answer [5]. Salomons et al. built on these findings and investigated if people will conform when there is no objective correct answer. They conducted an experiment in which participants were asked to play a game with three robots. The robots used in this experiment were embodied; they were given names and each had a unique voice and was uniquely dressed [4]. This contributes to the suggestion of personality and thus human-likeness. They found that participants who saw the robot's initial answers, changed their own final answers more and thus conformed more than participants who only saw everyone's final answers [4]. Their results show that in one third of the test rounds, people conformed to the group of robots, which is a similar outcome at the experiments done by Asch in the 1950s. Salomons et al. conclude that participants believe the robots may be better at the given task than they are [4]. It seems to be the case that this study shows conformity with robots because there was no objective correct answer to the questions. However, there are other aspects that could distinguish this study from previous studies, that failed to show conformity with robots. For example, the type of task.

It makes sense that the level of conformity is higher in ambiguous or unclear situations, when people are not fully convinced of their own beliefs. They are more likely to doubt themselves and accept influence from another person or group. The type of task and the level of clarity are different for humans and computers. Computers are known to be better at certain specific tasks than humans. The capabilities of both a human and a computer affects their credibility when it comes to certain tasks. Research shows people are more likely to trust a robot on analytical tasks, while they trust other humans more with social tasks. Nicholas Hertz studied people's willingness to consider advice from a non-human agent and to what extent this depends on the type of task given, and found that if the participants knew the type of task before choosing an agent, they chose machines more often than when they did not know the task beforehand. When given social tasks, participants more often chose a human as their advisor. He also found that participants conformed more strongly with the agents on social tasks as the advisor's human-likeness increased [10]. This indicated that in general, people are more likely to choose a human advisor. A similar experiment was conducted to test whether the degree of physical human-likeness affects conformity. The researchers found that human-likeness did not affect conformity. The results did show that people conformed more often with more ambiguous tasks [3]. When studying conformity with robots, the degree of human-likeness is not the only important factor to keep in mind. The type of task we trust another human to do better than ourselves, is not naturally the type of task a computer can do better too. Even if a robot seems more human-like because of its voice or appearance, there is still a clear-cut distinction between human and non-human. Not only human-likeness, but also the type of task seems to be an important factor in conformity. We should study social human groups.

D. Voice

Speech is an important element of social robots. Not only the presence of a voice, but also the way it sounds can influence the perception and interaction of humans with social and conversational robots.

Different social robots have different types of speech, like different degrees of human-likeness. Over time, voices of social robots have developed from unnatural and synthetic to more natural and humanlike. Cabral et al. studied the effect of synthetic voice on the evaluation of virtual characters in the context of audio-visual applications. They conducted an experiment to evaluate how synthetic speech impacts the perception of a virtual character and found that people rated a real human voice as more understandable, likeable and expressive than the synthetic voice used in the experiment. In two different conditions, they combined a human voice and a synthetic voice both with the same virtual character. Results show that the voices did not have a significant effect on the their ratings of the character's appeal, credibility and human-likeness [13]. Speech seems to be more important than visuals when people make judgments about understanding content delivered by the character [13][14]. This would be expected when the information is communication through speech. Cabral et al. focused on the evaluation of virtual characters, and similar results may be observed in human-robot-interaction. If the voice did not affect the human-likeness of the character, this is probably because of the visual image. When the visual image is taken away, the voice is the only factor to judge from and it might influence the credibility.

Perception of a voice is determined by the sound and style of speaking. This affects a human's cooperation with a robot. Different speaking styles of the same humanoid robots are preferred for different tasks. Goetz et al. [15] show that if a robot is speaking in a playful way, people are more willing to respond to the robot's instructions for a simple task, while they are more willing to respond to its instructions on a more serious task if the robot speaks in a serious way [7][15].

According to Christian Sandvig, there is a tension between the claim to objectivity and the performance of objectivity itself [16]. He argues that transparency and trust are processes that must be seen in order to be believed but the issue with algorithms is that for the most part they can't be seen [17][16]. A robot's performance of human sociality, specifically in the use of language, contributes to trust and persuasion.

The increasing presence and popularity of spoken language technology consumer products can be seen as an important step towards more advanced conversational agents. According to Roger K. Moore, the usage of these devices is surprisingly low. He suggests that this is partly because inappropriate humanlike voices of non-human agents might deceive users into overestimating their capabilities [18]. The humanlike voices allow users to have high expectations that the device cannot meet, which creates a conflict. Moore compares this to the uncanny valley theory: it results in the opposite of what was intended and therefore stands in the way of achieving acceptance.

For people without an extensive technological understanding about AI, it is almost impossible to know what they can expect from an intelligent agent. Making AI sound more human might make interaction with it more natural and seamless but it contributes to the conflict of expectations.

Moore highlights the benefits of giving them an appropriate voice. He argues that a more appropriate non-human voice would be one that is intelligible but robotic. Giving them a non-human voice instead of a human voice would help aligning the visual, vocal and behavioral affordances and thus the expectations it will create [18]. Expectations are easier to manage, especially for people without extensive technological understanding. It will remind naive users of the difference between humans and machines, and will make it easier for them to recognize limits of the robot's language capabilities.

While the sound of the voice may affect expectations, more aspects are relevant when researching human-robot-interaction. For example, other research shows there is no effect of human / synthetic voices on ratings of appeal, credibility and human-likeness [14]. The preferred type of voice might also depend on the type of task or the physical appearance of the robot. It is shown that the type of task determines which sound and speaking style is preferred from a robot [7][15]. The difference between experiments with embodied robots and experiments with disembodied robots is also relevant. If the robot has a body or another type of visual representation, the voice may not be aligned with the movements or behavior. However, if a robot is disembodied, this is less important. Recent development and increasing popularity of consumer-driven conversational assistants suggest that human voices benefits the interaction. Making AI sound more human is a logical step to simulating human appearance and behavior and with that, possibly, acceptability. Although some researchers are critical of the effects of giving conversational assistants a human voice, these developments need to be taken seriously in the rise of voice-enabled assistants.

E. Key Findings

In summary, conformity with social groups has been shown on the level of compliance, but in order to show a similar effect with non-human agents, some things need to be taken into account. Conformity with non-human agents is shown to be influenced by the degree of human likeness of the agent, but also by the type of task, the ambiguity of the task and the objectivity of the answers.

Conformity has not yet been shown in disembodied conversational agents. When human likeness in increased, it is often done so by improving the physical appearance or a combination of appearance and speech. However, in the light of contemporary developments in voice-enabled speech assistants, we are interested in the effect of speech on conformity. It has been shown that conformity is higher when tasks are more ambiguous, therefore we will study conformity in tasks with objectively correct or incorrect answers. We are primarily looking for differences between voices and speaking styles and how this can affect the interaction with the non-human agents. There are arguments why a synthetic voice would prevent conflicts of expectations and thus benefit the interaction, and arguments why simulating human speech would make the interaction more seamless and thus benefit the interaction. Measuring conformity with conversational assistants that have different voices and speaking styles can

give new insights in the effect of speech in human-robot-interaction and inspire further research.

III. METHOD

The question we aim to answer in this study is to what extent people conform to a human-like or robotic sounding AI voice assistant. Aside from two conditions with different voices, there is a third condition in which the assistant communicates through text. To measure the conformity rate in all three conditions, and find if there is a difference in conformity between the three groups, a qualitative research method is used.

Participants were randomly assigned to one of the three groups and asked to complete a general knowledge quiz, followed by a short survey about their demographic and conformity traits. We wanted to assess if people conform to a virtual assistant and if so, how often. We would also measure if there is a difference between text, a robotic voice and a human voice. In order to answer this question, we chose a quantitative approach. We divided the participants randomly into three groups of similar size.

A. Experiment

In order to answer the question to what extent people conform to a human-like or robotic sounding AI voice assistant when answering a series of multiple-choice questions under time pressure, we conducted an experiment. The participants (N = 162) were divided randomly into three groups. We aimed to get as many participants as possible and did not restrict them based on demographic characteristics. The participants were given a URL to a website with a short introduction and a start button that randomly directed them to one of the three quiz pages. The participants did not know there were different conditions and what the aim of the experiment was. Since the test was accessible online, all participants completed the test on their own device: a pc, a tablet or a smartphone. There was no controlled environment. We think it is important to study an interaction in the same way this would take place in the real world. People who interact with a conversational assistant in the real world, most likely do so on their own devices and in their own usual environment. If we would do the experiment in a controlled environment, the results would less likely be an accurate representation of interaction in the participants' daily lives.

We wanted to make sure people would not participate more than one time, since this would influence our results. To discourage people to do so, we mentioned on the introduction page that participants could only do the test once. However, it was possible to do the test multiple times. We gathered demographical data and combined this with other data such as participants' IP addresses. This allowed us to rule out double participations.

In two of the three quizzes, participants were asked to turn on the sound of their device before starting the quiz, to make sure they could hear the assistant. During the quiz, participants were asked to answer 20 general knowledge multiple-choice questions within 30 seconds. There was one additional example question at the start of the quiz to illustrate the process of answering the questions twice. Participants had two chances to answer every question, because we wanted to know whether they would change their initial answer based on the assistant's advice. The questions were either displayed on the screen or was read out loud by the assistant. When they submitted their answer or the 30 seconds were over, they got the same question again with their own answer still checked. We programmed the assistant to share either factual information with a different degree of usefulness, or share its thoughts. It did not even give factually incorrect information, only selective and sometimes misleading information.

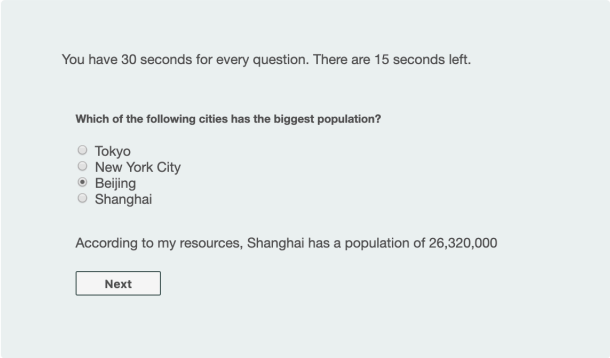

Fig 1. A screen capture of the quiz in the 'text' group. Screen capture taken after the first answer was given.

The questions did not have one obviously correct answer: at least two of the answers were plausible so when the AI advised to choose an incorrect answer, the risk that participants distrusted the AI because of it was limited. In some cases, the assistant would not suggest another answer but would merely ask the participant if they were sure or tell them the given answer was probably incorrect. The assistant would try to influence the participants regardless of whether they answered the question correctly. The advice from the assistant was based on the initial answer of the participant. The assistant would always either give vague and useless information, or try to steer the participant in the direction of another, specific answer. It would never confirm the initial answer given by the participant.

The participants would also not immediately get the answer since this could influence their trust in the assistant. This experiment measures how often participants changed their answer to conform with the AI. We measured if and how often participants conform to the AI and if there is a difference between the groups.

B. Conformity trait scale

After the quiz, a short number of demographical questions and a 10 item conformity scale followed. The conformity scale was used to assess the conformity tendency of all participants. It does not show why people conform more than other people, only how much people report to conform in daily-life situations. We want to know how much of our participants have a high tendency to conform and if this influences the results of the experiments. We used the conformity scale based

TABLE I: CONFORMITY SCALE ITEMS AND RELATIONSHIP

"I often rely and act upon the advice of others"	+
"I would seldom change my opinion in a heated argument on a controversial topic"	-
"Generally, I'd rather give in and go along with majority of others for consistency"	+
"Basically, people around me are the ones who decide what we do together"	+
"Environmental information can easily influence and change my ideas "	+
"I am more independent than conforming in my ways"	-
"If someone is very persuasive, I tend to change my opinion and go along with them"	+
"I don't give in to others easily"	-
"I tend to rely on others when I have to make an important decision quickly"	+
"I prefer to make my own way in life"	-

on Mehrabian and Stefl [19]. Each item was scored ranging from −2 strongly disagree to +2 strongly agree. Six items were positively scored and four items negatively. One item from the original scale, concerning family tradition and political decisions, was excluded [19]. The items are shown in Table I.

C. Materials

Due to time constraints and technical limitations, we used a 'Wizard-of-Oz' approach: subjects interacted with the computer system that they believed to be autonomous, but which was actually pre-programmed by the researchers. There was no actual intelligent system. Instead, we programmed the messages and answers of the assistant into the quiz. Before starting the quiz, the assistant introduced itself through either displayed text or audio. The assistant communicated the same information in all three conditions, only in the last two audio recordings were played.

The robotic voice was generated with the use of text-to-speech software. The human voice was recorded by a voice actor as a list of sentences. The actor did not know the questions and correct answers they belonged to and had no understanding of the content and context. She took a pause between every sentence, to make sure they were pronounced as neutral as possible. In this way we tried to make sure there were no social signals directing people to a specific answer. To make it sound more human, she sometimes included speech disfluencies and conversation fillers such as "oh" or "uhm".

The sentences recorded were the same in both groups, but sounded less natural in the robot voice than the human voice.

D. Participants

The test group consisted of 162 people, between the ages of 17 to 61. Participants were recruited through social channels and email. We aimed to get as many participants as possible and did not restrict them based on demographic characteristics. We did however keep track of demographic characteristics such as gender, age, nationality and relevant technological knowledge since they could be considered extraneous variables. The test was accessible online: participants were given a URL to a website which directed randomly to one of the three conditions.

E. Data analysis

We will analyze the results with a one-way between groups analysis of variance (ANOVA) to investigate the impact of the assistant's mode of communication towards conformity. This is the most suitable statistical test when there are more than two independent groups. We are working with one independent variable that has three possible values (text, robotic voice and human voice), and the percentage of times participants conformed as the dependent variable.

This test compares the means of the three groups and shows whether one of the three is significantly different from the others. In this event a post hoc test shows where the difference lies.

IV. RESULTS

The quiz results and demographic survey results were analyzed to test if people conform more when the assistant has a voice and whether there is a change in conformity when the assistant sounds human rather than robotic. We conducted the experiment online to increase the chances of getting more participants in a short amount of time (N = 162).

The first group (text) consisted of 55 participants, the second group (robotic voice) of 54 and the third group (human voice) also of 54. We aimed to get equally sized groups. However, we did not have full control over this because we used the random function in our program.

A. ANOVA

The quiz contained 20 questions but not all participants answered all questions. We took the number of times they changed their answer, correctly or incorrectly, and divided it by the number of questions answered to get the percentage of conformity. If they answered less than half of the questions, their results are left out. This was the case for 2 of the total 164 that participated. These cases are not included in the *N*.

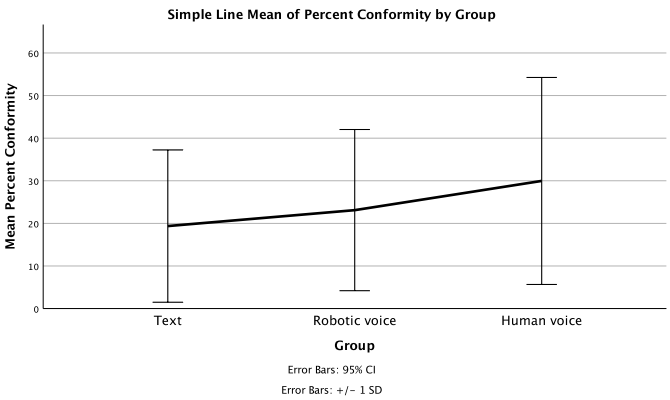

Fig. 2: Means plot of Percent conformity for all three conditions.

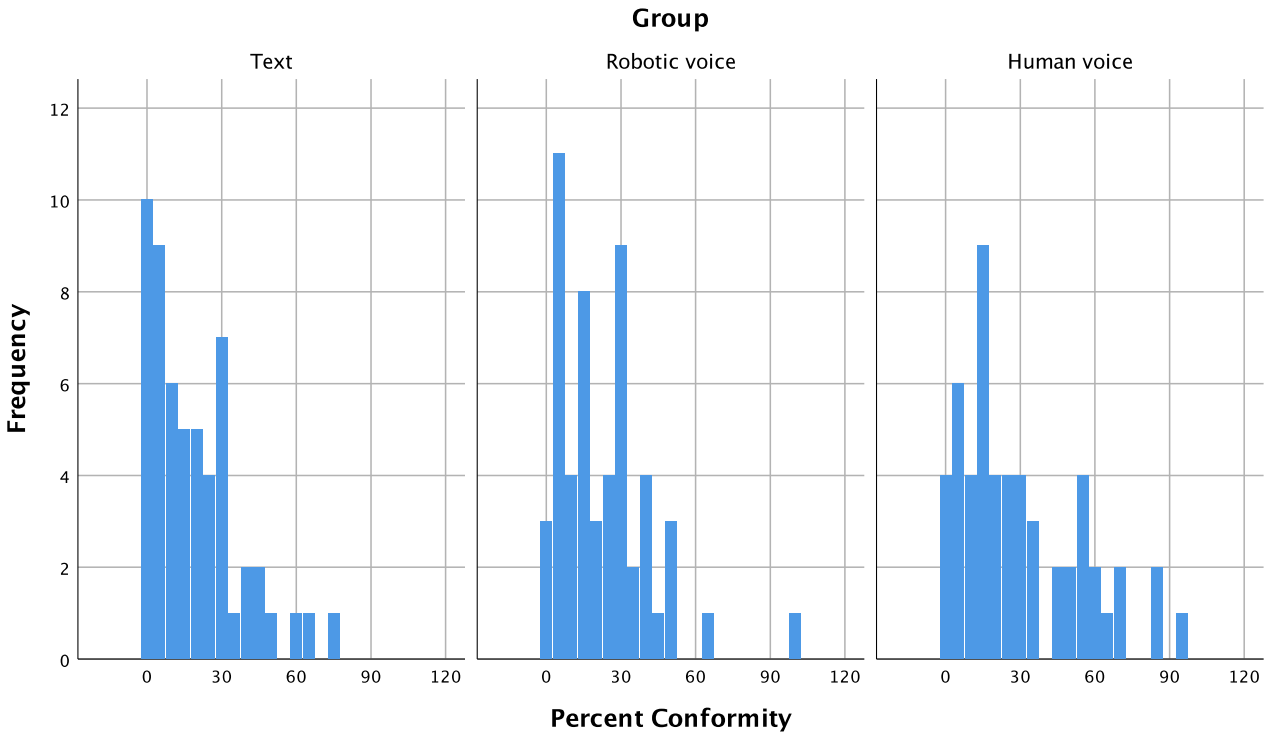

Fig. 3: Frequency histogram for all three condition

A one-way ANOVA was used to compare the means of the three groups. The means are plotted in Fig2. The frequency histograms are shown in Fig 3.

The results show a significant effect of mode of communication on conformity at the $p < .05$ level for three conditions $F(2, 160) = 5.14$, $p = .026$ (Table II). We did not know beforehand which group was most likely to be different, so we did not specify a priori contrasts.

Because there is a significant result, a Tukey HSD (using an alpha of .05) post hoc analysis was done (Table III) This revealed a significant difference between text ($M = 19.36$, $SD = 17.87$) and a human voice ($M = 29.96$, $SD = 24.27$).

Participants conformed more when the assistant sounded like a human, than when the assistant only communicated through text. However, there was no significant result between the robotic voice ($M = 13.11$, $SD = 18.92$) and text ($p = .608$), nor between the robotic voice and the human- voice ($p = .196$). 89% of participants changed their answer at least once.

TABLE II: RESULTS OF THE ONE-WAY ANOVA TEST SHOWING THERE IS A SIGNIFICANT RESULT BETWEEN GROUPS.

ANOVA					
Percent conformity	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3142.799	2	1571.399	3725	.026
Within Groups	67495.987	160	421.850		
Total	70638.785	162			

The results of the Shapiro-Wilk test were significant based on an alpha value of 0.05, $W = 0$. $p < .001$. This result suggests the residuals of the model are unlikely to have been produced by a normal distribution, indicating the normality assumption is violated. However, with large sample sizes (> 30 - 40), the violation of the normality assumption should not cause major problems. According to Ghasemi and Zahediasl, this implies that we can use parametric procedures, even when the data are not normally distributed [20].

A linear regression analysis was conducted to assess whether gender, age, Technological knowledge, Conformity trait score and Group significantly predicted Percent conformity. This showed a significant result for Group and for Technological Knowledge ($p = .045$). This implies that technological knowledge can account for variance in Percent conformity beyond that which can also be predicted by non-significant variables such as gender, age and the Conformity Trait score.

Participants who reported to have technological knowledge about AI conformed less than people who reported to not have relevant technological knowledge. We split the data set into two groups, high technological knowledge ($N = 32$) and low technological knowledge ($N = 119$). 11 participants did not answer the question about technological knowledge. We ran the same ANOVA again on both data sets. This shows a significant difference between text and human voice in the low technological knowledge group ($p = .047$). In the group with high technological knowledge, no significant difference was shown.

TABLE III: RESULTS OF THE TUKEY HSD POST-HOC TEST SHOWING THE DIFFERENCES BETWEEN ALL GROUPS.

Multiple Comparisons						
Dependent Variable: Percent conformity						
Tukey HSD						
					95% Confidence Interval	
(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	Lower Bound	Upper Bound
Text	Robotic voice	-3.747	3.935	.608	-13.06	5.56
	Human voice	-10.599*	3.935	.021	-19.91	-1.29
Robotic voice	Text	3.747	3.935	.608	-5.56	13.06
	Human voice	-6.852	3.935	.196	-16.20	2.50
Human voice	Text	10.599*	3.935	.021	1.29	19.91
	Robotic voice	6.852	3.935	.196	-2.50	16.20
*. The mean difference is significant at the 0.05 level.						

No relation was found between the conformity trait and the number of times the participant conformed. This is noteworthy, given that the conformity scale is used to gain insight in how likely people are to conform in general. It would be easier to explain if there was a correlation between conformity score and the number of times participants conformed. However, it seems that the conformity scale does not predict conforming behaviour as measured in our experiment.

The statistical power of the experiment is 0.81. The results would be more powerful with a larger sample. We conducted the experiment online to get more participants but a trade-off is that we were not there when they completed the quiz. Although we took measures to limit the risk that they cheated, there was still the possibility that they cheated by doing the quiz together or looking up the correct answers.

V. Discussion

Overall, people conformed to the assistant to some extent in all conditions. This indicates that people conform not only to a group but also to a single agent, and not only to a human but also to a computer. This shows that conformity can occur with disembodied conversational agents, not just with embodied robots or conversational agents that have been given some kind of visual representation.

The results show a significant difference in conformity between text and human voice. Participants who were assisted by an agent with a human voice conformed more often than participants who were assisted by an agent who did not have a voice at all and only communicated through text. Reasons for participants to change their answers could include social pressure or a belief that the robot is (more) intelligent and probably knows more than they do. Because they conformed more to the agent with the human voice, this suggest that the participants had more trust in this agent and are more willing to accept advice from it.

While in the original experiments by Asch there was compliance with a group, the type of conformity that occurred in this experiment is different. Participants knew that they were communicating with an AI in all cases, so they were never under the impression they were communicating with another human being. It is unlikely that participants complied with the AI in the sense of Kelman's definition [11]. People do not desire approval from a computer at the same level they desire approval from (groups of) other people. We think one of the biggest factors that influenced people is the credibility of the source. It is shown that computers are better at specific tasks than human, especially tasks with low ambiguity [3][4][10]. We gave the participants answers to which one answer was objectively correct, meaning the tasks had low ambiguity. If participants believed that the AI had access to resources, for example through an internet connection, it is likely they perceived the shared information as credible.

This would be a case of internalization, if the motivation for conformity was the credibility of the source. There is less emotional motivation, like desire for approval or fear of being rejected by the AI. Participants had to internally believe the information given by the AI in order for them to conform with it. This would explain why participants conformed to the AI in general.

There is a significant relation between technological knowledge and conformity with the assistant. An explanation could be that people with more technological knowledge had a better understanding of how intelligent assistants (might) work and are more aware of the limitations of such systems, while people with less technological knowledge are more vulnerable to the influence of a so-called 'intelligent' system. People with high technological knowledge might be more skeptical towards the capabilities and credibility of the assistant than people with low technological knowledge.

While in general there is less emotional motivation to conform with a computer than with other people, there is a difference between the mode of communication of the AI. Participants in the human-voice group conformed significantly more than participants in the text group. Previous research has not yet formulated a theory on why people would conform more with a human-sounding robot that with a robotic-sounding robot. However, we think this could be caused by a form of identification: a type of conformity that occurs when an individual accepts influence from someone who is liked and respected. This is usually someone like a celebrity or a family member. However, it could be argued that when participants heard a human voice, they felt some kind of connection on a level that participants who read text displayed on their screen did not. It is plausible that people identify more with the human voice than a robotic voice because the human voice sounds more like them. More research is needed to study the psychological processes behind this.

The results also show that people are willing to accept the virtual assistant as intelligent enough to affect their own decision-making. Previous studies showed that people are more likely to conform if there is no objective correct or incorrect answers [5]. Other studies show that people conform more as the ambiguity of the task increases [4]. The nature of the task determines to a large extent whether or not people accept a machine as their advisor [10]. All questions in our quiz were general knowledge questions with an objective correct answer. If participants believed the assistant had access to resources

and gave them the most relevant information to answer the question, the task of the assistant would be considered an analytical one rather than a social one. The fact that they accepted the assistant enough to conform with it, is in line with earlier research that showed people are willing to accept the advice of a machine if it concerns a task with low ambiguity [10].

When a virtual assistant has a voice, it is more human-like than when it can only display messages in text. Speech is characteristic of humans, not of machines. Speech allows for a more natural and more seamless interaction with it, or it can be used to deliberately increase human-likeness of the AI agent.

While people conformed slightly more to the human voice than to the robotic voice, (Table III) the results do not show a significant difference between the two voices, indicating that the sound of the voice and the natural or synthetic way of speaking do not affect conformity. This suggests that people do not have significantly higher expectations of a human-sounding agent than a robotic sounding one. According to Moore, conversational agents with humanlike voices deceive users into overestimating their capabilities. The results of this experiment did not show that was the case specifically with humanlike voices rather than robotic voices. Nor do we see the opposite effect: the results do not show an uncanny valley effect where increasing the human likeness leads to a point where it evokes eeriness or revulsion.

The virtual assistants did actively deceive the participants into thinking they knew more and were more capable of answering the general knowledge questions correctly, while the information given by the assistants was often misleading.

If the results proved a difference between the human and synthetic voice, this would support Moore's claim that an appropriate voice has a beneficial effect on the interaction with these agents [18]. Nevertheless, a non-significant result does not reject Moore's claim.

We measured the level of conformity in the light of acceptability and trust within Human Robot Interaction. A higher level of acceptance and trust will lead to more conformity. Whether this is a positive or negative development depends on the field from which it is studied.

VI. Conclusion

The first hypothesis was that people conform to a single virtual intelligent assistant. We used the text condition as a base line to assess if people would conform at all in the quiz setting we used in the experiment. Overall, people conformed in around a quarter of the cases and 89% of participants conformed at least once, showing that there is a substantial degree of conformity. There is a significant difference between the text condition (base-line) and the human voice condition. We conclude that the first hypothesis can be accepted. Whether conformity is a beneficial effect in general, goes beyond the scope of this research.

The second hypothesis was that there is a difference between conformity with a voice-assistant and conformity with a virtual assistant that does not have a voice. This hypothesis can also be accepted because participants conformed more in the human voice group than in the text. We can conclude that adding a voice to a virtual intelligent assistant, benefits the interaction in terms of conformity.

The results also show a difference in conformity between the group with the robotic voice and synthetic way of speaking, and the group with a virtual assistant with a human voice. While the difference between these two groups is non-significant, in one of them participants conformed significantly more than in the text group, and in the other people did not. The human voice group is the only group in which people conformed significantly more than the group we used as a base line. This indicates that the robotic voice group and the human voice differ.

The experiment and results were discussed in relation to Asch's conformity experiments and other previous research to place conformity with non-human in a broader perspective.

In light of our increasing reliance on AI assistants, this study investigated the influence of voice on interaction with these assistants. It does not suggest that developing virtual personal assistants to sound more human pays off in terms of acceptability. While some researchers argue that a more synthetic voice is needed to avoid a conflict of expectations [18], this study does not show a need to move away from human voices either. Moving from text to sound made a significant difference, one that is not made by developing speech to be more human.

References

[1] Leviathan, Y., & Matias, Y. (2018, May 08). Google Duplex: An AI System for Accomplishing Real-World Tasks Over the Phone. Retrieved June 21, 2019, fromhttps://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html

[2] Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field]. IEEE Robotics & Automation Magazine, 19(2), 98-100.

[3] Hertz, N., & Wiese, E. (2016, September). Influence of agent type and task ambiguity on conformity in social decision making. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 60, No. 1, pp. 313-317). Sage CA: Los Angeles, CA: SAGE Publications.

[4] Salomons, N., van der Linden, M., Strohkorb Sebo, S., & Scassellati, B. (2018, February). Humans conform to robots:Disambiguating trust, truth, and conformity. In Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (pp. 187-195). ACM.

[5] Brandstetter, J., Rácz, P., Beckner, C., Sandoval, E. B., Hay, J., & Bartneck, C. (2014, September). A peer pressure experiment: Recreation of the Asch conformity experiment with robots. In 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems (pp. 1335-1340). IEEE.

[6] Asch, Solomon E. 1956. Studies of independence and conformity: I. A minority of one against a unanimous majority. Psychological monographs: General and applied 70, 9 (1956), 1.

[7] Markowitz, J. (2017). Speech and language for acceptance of social robots: An overview. Voice Interaction Design, 2, 1-11.

[8] Mitchell, W. J., Szerszen Sr, K. A., Lu, A. S., Schermerhorn, P. W., Scheutz, M., & MacDorman, K. F. (2011). A mismatch in the human realism of face and voice produces an uncanny valley. i-Perception, 2(1), 10-12.

[9] Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. Cognition, 125(1), 125-130.

[10] Hertz, N. (2018). Non-Human Factors: Exploring Conformity and Compliance with Non-Human Agents (Doctoral dissertation, George Mason University).
[11] Kelman, H. C. (1958). Compliance, identification, and internalization three processes of attitude change. Journal of conflict resolution, 2(1), 51-60.
[12] Gerard, H. B., Wilhelmy, R. A., & Conolley, E. S. (1968). Conformity and group size. Journal of Personality and Social Psychology, 8(1p1), 79.
[13] Cabral, J. P., Cowan, B. R., Zibrek, K., & McDonnell, R. (2017). The Influence of Synthetic Voice on the Evaluation of a Virtual Character. In INTERSPEECH (pp. 229-233).
[14] Gong, L., & Nass, C. (2007). When a talking-face computer agent is half-human and half-humanoid: Human identity and consistency preference. Human communication research, 33(2), 163-193.
[15] Goetz, J., Kiesler, S., & Powers, A. (2003, October). Matching robot appearance and behavior to tasks to improve human-robot cooperation. In The 12th IEEE International Workshop on Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003. (pp. 55-60).
[16] Sandvig, Christian. "Seeing the Sort: The Aesthetic and Industrial
Defense of "The Algorithm." Media-N 10.3(2014).
[17] Phan, T. (2017). The Materiality of the Digital and the Gendered Voice of Siri. Transformations (14443775), (29).
[18] Moore, R. K. (2017, August). Appropriate Voices for Artefacts: Some Key Insights. In 1st International Workshop on Vocal Interactivity in-and-between Humans, Animals and Robots.
[19] Mehrabian, A., Stefl, C.A., 1995. Basic temperament components of loneliness, shyness, and conformity. Social Behavior and Personality 23, 253–264.
[20] Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: a guide for non-statisticians. International journal of endocrinology and metabolism, 10(2), 486.