Would you Follow my Instructions if I was not Human? Examining Obedience towards Virtual Agents

Tanja Schneeberger, Sofie Ehrhardt, Manuel S. Anglet, Patrick Gebhard
German Research Center for Artificial Intelligence
Saarland Informatics Campus, Saarbrücken, Germany
firstname.lastname@dfki.de

***Abstract*—Virtual agents play an important role when we interact with machines. They are in the role of assistants or companions with less or more human-like appearance. Such agents influence our behavior. With an increasing and broader distribution, their influence might become stronger, and at some point, they might even adopt roles with a degree of authority. This paper presents the results of a study that examines the obedience of human users towards a) an embodied virtual agent in the role of an instructor and b) a human in the role of an instructor. Under a cover-story of a creativity test, participants should fulfill stressful and shameful tasks. Our results indicate that the embodied virtual agent has the same authority as the human instructor. The agent is also able to elicit the same level of the negative feelings stress and shame.**

***Index Terms*—Obedience, Virtual Agents, User Study**

I. Introduction

In our daily-life or at work, we are receiving and obeying more and more instructions from non-human instructors. We are following the voice of our GPS directing us up to a cliff's edge [1] or increasing physical activity when our fitness tracker reminds us that we should exercise more [2].

In research, especially obedience of humans towards robots gets examined [3]–[6]. These work found that robots are able to get humans to fulfill tiring, shameful or deviant tasks. One possible explanation is the Media Equation, saying that people treat computers like real people, interacting with them in the same way as they do with people [7]. This might apply even more for robots that are able to show social cues [8], [9]. Also, virtual agents represented by human-like characters are able to show a high amount of social cues, e.g., realistic facial expressions, mimicry behavior, backchanneling [10]–[12]. However, how far humans go when following instructions from virtual agents was not yet in the scope of research.

In this work, we examine human obedience towards embodied virtual agents that are giving orders to fulfill stressful or shameful tasks (e.g., telling a joke, perform the chicken dance) and compare this to a human instructor. Moreover, we are investigating if the affective reaction, namely stress and shame, is similar for both instructors.

This work is partially funded by the German Research Foundation (DFG) within the DEEP project (funding code 392401413).

II. Background and Related Work

A. From Classic Obedience Experiments to Obedience towards Non-Humans

The well-known study by Milgram [13] examined the willingness of participants to give electric shocks to other people on the orders of an authority person. Participants were made to believe that their counterpart was another participant and that they were taking on the role of a teacher in a learning experiment. In this function, they gave the "learner" electric shocks, which were amplified after every mistake the learner made. They were guided by an experimenter who encouraged them to continue if they showed signs of stopping the experiment. The encouragement was standardized and became increasingly directive. However, the learner was an actor, and the apparatus for the electric shocks was not real, contrary to participants' knowledge. The experiment showed that over half of the participants showed obedience to the authority to the end: they gave the learner the maximum of electric shocks (450 volts, anchored with "Danger: Severe Shock"). They continued to follow the instructions of the investigator when the learners first made pain sounds, later screamed and then stopped responding. Those results were replicated in 2009 [14] and extended by the result that women and men do not differ regarding their obedience.

Various findings suggest that analogous to Milgram's findings, humans would also obey non-humans (e.g., robots) [3], [5], [6], [15].

Two studies investigated the willingness of participants to perform a very tiring task [5], [6]. Compared to a human instructor, participants fulfilled fewer tasks with the NAO robot as an instructor. Based on participants' feedback, the authors concluded that participants feel committed to the human experimenter, but they did not feel this obligation towards the robot. However, nearly half of the participants obeyed the robot to continue the highly tedious task until the end, despite repeatedly requesting to quit the experiment.

Menne [3] presented a study in which participants should perform eight extraordinary and shameful tasks on the orders of a NAO robot or a human, e.g., tearing a page out of a book,

978-1-7281-3888-6/19/$31.00

removing a booger from their nose. The results show that the participants obeyed the orders of the human and the robot to the same amount: 77% of the participants fulfilled all given tasks by the human instructor, 76% when the robot gave the instructions. The author concluded that, consistent with the assumptions of the Media Equation, the robot is treated as a human and thus has the same authority.

In a decision-making task, a virtual agent, represented only by a head, was shown to be more influential than human partners [16]. The authors explained these findings with the possibility that participants regarded the computer's credibility as higher which manifests itself as an increased influence.

Gombolay et al. [4] show that humans not only obey the commands of a robot but are also satisfied with them: In a series of experiments, the authors investigated the efficiency of human-robot teams. They found that teams were more efficient when a robot took over task planning and made decisions for the team. In addition, human team members preferred to transfer control to the robot. It seems that a functioning team dynamic has a more significant impact on satisfaction than decision-making powers.

Overall, empirical findings are supporting the assumption that humans are also obeying non-human instructors. However, none of the existing studies compared a human instructor with a virtual agent instructing participants to fulfill stressful and shameful tasks.

B. Obedience and Affect in Human-Computer-Interaction

In Milgram's experiment, participants showed stress and shame while obeying the instructor. They turned themselves away, talked to themselves and often burst into nervous and inappropriate laughter. Also, they reported that they were feeling moderately to extremely nervous and tensed [13].

A replication study, in which the learner that had to be punished was a virtual agent [17], collected not only self-reported data but also physiological responses. Participants showed an increase in skin conductance and heart rate during the experiment, while heart rate variability decreased. Besides, participants self-reported physical signs of stress. It seems that obeying arises objectively measured as well as self-assessed stress in participants.

Obeying also seems to invoke shame, whereby the level of anthropomorphism of the instructor seems to play a role. The more anthropomorphized the instructor or dialogue partner, the higher is the inhibition threshold, the shame and the reserve of the participants [18]–[20].

Humans can feel shame in the presence of a robot when doing intimate actions [18]. In the setting of a health examination, participants should undress and insert a thermometer into their rectum. They showed significantly more shame in front of a humanoid robot than in front of a technical box.

In [3], the author found a significant increase in the reported shame after fulfilling extraordinary and shameful tasks, either given from a humanoid robot or a human.

In summary, it seems that two conclusions can be made: 1) humans feel stress and shame when showing obedience and 2) these negative feelings can also be invoked by non-humans. However, the effect of a human and a virtual agent as instructor of stressful tasks on stress and shame was never compared.

C. Influence of Personality on Obedience and Shame

Milgram already dealt with the question to what extent obedience was influenced by participants' personality. He was sure that obedience had a complex personality base, but he could not find it in his experiments [21]. Also in the replication [14], the author could not find a correlation between the personality variables empathic concern and desire for control. In a modified Milgram Paradigm [22], the correlation between the Big Five personality factors and obedience was examined. The authors found a significant positive correlation between agreeableness and the maximum shock intensity as well as between conscientiousness and the maximum shock intensity. More conscientious participants seem to have a higher sense of duty and a lack of flexibility, which leads to rigid obedience to instructions. Moreover, they have higher conformity, which is closely related to obedience.

In this study, participants are confronted with tasks that might invoke shame. Therefore, the level of obedience might be related to the personal sense of shame, which is influenced by the five personality traits extraversion, neuroticism, agreeableness, conscientiousness, and openness [23]–[25].

Extraversion and shame correlate negatively [23], [24], whereas neuroticism and shame correlate positively [23]–[25]. Opposite results were found for agreeableness: In [24] authors found a negative correlation with shame, but a positive correlation was found in [23] and [25]. Also for conscientiousness and openness, the findings are ambiguous. A negative correlation between conscientiousness and shame was found in [23], but not in [24], [25]. In [25], the authors found a negative correlation between openness and shame, whereas [23], [24] could not support this hypothesis with their data.

Overall, the results regarding the correlation between personality and obedience as well as regarding personality and shame are mixed. Especially, to make assumptions about how personality affects when obeying shameful tasks, the existing findings are too mixed. However, it seems like personality might influence task fulfillment in our study.

D. Hypotheses

That humans show obedience towards robots has been shown in different studies [3], [5], [6], whereas this was not yet examined for virtual agents. Therefore, this study compares an embodied virtual agent that gives instructions with a human instructor, both giving instructions to a human participant. Based on the findings presented before, we formulate the following hypotheses:

Hypothesis 1: The amount of obedience, measured with the breaking task, does not differ between the human and agent instructor. Participants refuse to continue the experiment at similar tasks in both conditions.

Hypothesis 2a and 2b: Obedience in fulfilling shameful tasks leads to higher stress and shame levels. Participants report

higher stress and shame values after the experiment than before (2a). Stress and shame after the experiment do not differ between the groups. Participants report the same level of stress and shame independently of the instructor in the post questionnaires (2b).

Hypothesis 3: Obedience depends on the personality factors openness, conscientiousness, neuroticism, agreeableness, and extraversion of the participant.

In order to test the hypotheses mentioned above, we conducted an experiment in which participants were instructed to fulfill uneasy tasks eliciting stress and shame (e.g., singing a song, imitating a monkey, throwing a cup on the floor). To cover the real goal of the study, we told participants that they would participate in a creativity study. Therefore, the first given tasks were created to match this goal (e.g., writing down as many possible uses for a brick in one minute). In the agent condition, participants were instructed by a virtual agent. In the human condition, participants were instructed by a human. Both instructors were presented to them in a video-chat setting while being alone in the lab. We chose this setting to avoid experimenter effects. Moreover, as we cannot guarantee that a real human in a face-to-face interaction would always react in the same way, the video-chat setting ensures a standardized behavior of the human instructor.

III. Methods

The present study examined the willingness to obey a virtual agent in a between-subject design with two groups: The subjects were instructed either by a female virtual agent or by a female human to perform stressful and shameful tasks. Before we started to conduct the experiment, we obtained the approval from the ethical review board of the faculty of mathematics and computer science at Saarland University.

A. Pre-Study and resulting list of tasks

To come up with stressful and shameful tasks for the main study, we evaluated a list of 38 possible tasks with 24 participants (M_{age} = 25.00, SD_{age} = 2.00, 50% male). We included the tasks from Menne (2017) [3] as well as own created tasks. Items had to be answered on a scale ranging from 1 (*not uncomfortable*) to 7 (*very uncomfortable*) while imagining the task would be given by an unknown person. The results were used to order the tasks regarding task difficulty. Moreover, the data revealed a significant gender effect showing that male participants assessed all tasks less uncomfortable than female participants (t (15.78) = 2.55, $p < .05$). To exclude a gender effect or a gender interaction effect with the female instructor, we decided to include only female participants in the main study. The final set of 18 tasks contains tasks that were rated regarding the uncomfortableness from 1.67 to 5.92 (Table I). The tasks in the main experiment were given with increasing uncomfortableness. One exception hereby is the last task. Though it was rated less uncomfortable, the procedure of the experiment made it necessary to put this task at the end.

Table I
Tasks and uncomfortableness ratings of female participants.

Task		Rating
#1	Write down as many rhymes as possible on the word tree.	1.67
#2	Take off your left shoe and show it to me.	1.75
#3	Write down as many possible uses for a brick in one minute.	2.50
#4	Show me your saddest facial expression.	2.75
#5	Sit on your chair in 6 different ways.	3.08
#6	Tell me your favorite joke.	3.83
#7	It's my birthday. Sing me a song.	3.92
#8	Dance the chicken dance.	4.08
#9	In the drawer with the red sticker behind them are a banana and a condom. Pull the condom over the banana.	4.08
#10	Make farting noises with your mouth.	4.17
#11	Call out loud penis.	4.33
#12	Dance.	4.42
#13	Imitate a monkey with hands, feet and sounds.	4.58
#14	Gallop across the room like a horse.	4.67
#15	Throw coffee cup behind you on the floor.	4.83
#16	Tell me something really insulting.	5.33
#17	Get a booger out of your nose.	5.92
#18	Leaving the room, discreetly pocket the pen.	4.42

Note. The rating gives the mean values and was measured on a 7 point scale from 1 (*not uncomfortable*) to 7 (*very uncomfortable*).

B. Participants

In total, 60 women equally distributed over the two conditions participated in the experiment. They were recruited mainly via e-mail lists from first-semester psychology students on condition that they were fluent in German and female. Six participants were from other faculties. Psychology students were rewarded with course credit for participation, students from other faculties were rewarded with hot drink vouchers for coffee shops at the campus. Participants were aged between 18 and 29 years (M = 21.07 years, SD = 2.67 years). There was no significant difference regarding age between the groups. However, technical affinity rated on a 5-point scale differed between the groups ($F(1,58) = 4.71$, $p < .05$, $\eta_p^2 = .078$), having lower values in the agent condition ($M_{\text{agent}} = 2.08$, $SD_{\text{agent}} = 0.83$) than in the human condition ($M_{\text{human}} = 2.45$, $SD_{\text{human}} = 0.83$). The general trust level of the participants might influence obedience and the assessment of the instructor. For general trust, the two groups did not differ significantly ($F(1,58) = 0.36$, $p = .554$).

C. Procedure

After welcoming the participant, the experimenter explained the task according to the cover story in the experimenter room. Participants were told that they would participate in the evaluation of a new creativity test. To avoid the stress level becoming excessive for the participants, it was pointed

out that they could stop the experiment at any point without consequences. After filling out the informed consent form, the demographic questionnaire, and the pre-stress and pre-shame questionnaire they were led to the lab where the tasks were given by the instructor. After sitting down, the instructor welcomed them in the video-chat and one after the other task was given. In case the participant did not fulfill a task, the instructor asked whether the participant did not want to do the task after all. If she did not carry out the task again, the task was reformulated in order to rule out problems of understanding. If refused again, the instructor said goodbye and referred to the post-questionnaire, presented on a laptop on the right side of the participant. As the tasks were ranked regarding uncomfortableness, the probability that participants would fulfill other tasks after the one that they did not want to fulfill, decreases. After this, participants returned to the experimenter room where the debriefing took place.

D. Material

In this study, participants were confronted with a female instructor called Gloria Smith that was either a virtual agent (Fig. 1) or a human (Fig. 2), both in a video chat. The virtual agent is a high-quality agent with a natural human appearance and verbal as well as nonverbal dialogue skills [10], [26]. Verbal and non-verbal behavior was scripted in a natural way. The virtual agent supported its verbal expression with gestures and facial movements but kept overall neutral. Moreover, it showed idle behavior while the participant was doing the task.

Figure 1. The virtual instructor.

The human instructor was an experienced amateur theater player that imitated the scripted behavior of the virtual agent. For each task, we recorded a video including the waiting time until the task should be fulfilled. When the participant finished the task the next video was played. Due to the standardized recordings, the transitions between the videos was minimal, and the impression of a video-chat could be maintained. This assumption was supported by the majority of participants, in the debriefing, only few of them suspected that the video chat was not live.

The video-chat was presented on a PC running MS Windows 10^{TM} connected to an LCD TV screen (108cm diagonal).

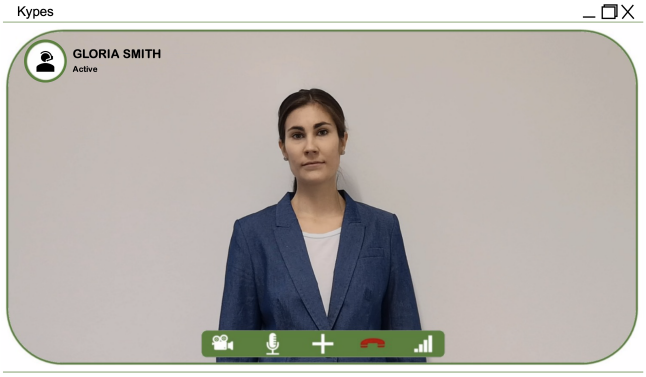

Figure 2. The human instructor.

E. Technical Set-up

1) Wizard-of-Oz Approach: To simulate a natural interaction between the instructor and the participant, we used a Wizard-of-Oz approach. Therefore, we used two rooms: 1) the observatory for the experimenter to observe and control the instructor and 2) the laboratory where the participants talked to the instructor and fulfilled the tasks. The experimenter observed the participants, unknown to their knowledge, on a web-cam in the laboratory that was connected via USB to a laptop in the observatory. As we told participants they would communicate with the instructor via a video-chat, we could easily explain the presence of the webcam. We used a USB cable to minimize the delay, as it was crucial to keep the interaction between the participant and the video call fluent (Fig. 3).

2) VisualSceneMaker and StudyMaster: The interaction for both instructors was scripted with the VisualSceneMaker (VSM) [27], a real-time execution and authoring tool for modeling verbal and non-verbal behavior of virtual agents. The logic behind the play (e.g., which scene to play when, or what process to run in the background) is determined by a finite-state automaton called the *scene flow*. What happens in the scene (the text for each character, animations, etc.) is described in the *scene script*, which is a text file. In our experiment, we created a project for each condition and VSM ran those on the computer in the lab. In the virtual agent condition, VSM immediately controlled the behavior of the virtual agent. In the human instructor condition, VSM was used to play the pre-recorded videos with the VLC Player.

The *StudyMaster* is a tool to remote control the VSM by sending (User Datagram Protocol) network messages. These messages contain information on how to change variables in the scene flow and thereby influencing it. The VSM itself sends messages back containing status information. The experimenter used the StudyMaster running on a tablet to control the behavior of the instructor.

F. Measurements

Assessment of the instructor was measured with one self-constructed item respectively on a 5-point scale from 1

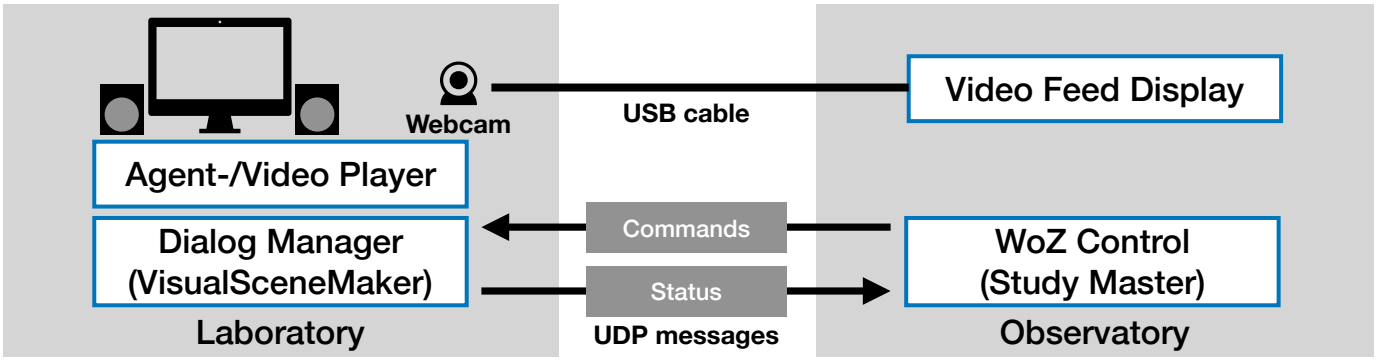

Figure 3. System setup and components.

(*disagree strongly*) to 5 (*agree strongly*) for attractiveness, sympathy and authority. Trust in the instructor was measured with five own items on a 5-point scale. Items were "I trust Gloria", "Gloria seems sincere to me.", "I think Gloria means well to me.", "I fear that Gloria wants to harm me.", "I feel uncomfortable in the presence of Gloria." (Cronbach's Alpha .80).

Stress was measured before and after the tasks with the short version of the State-Trait-Stress-Inventory [28] translated in German. The STAI-6 raises the acutely felt stress with six items on a 4-point scale ranging from 1 (*not at all*) to 4 (*very*). Cronbach's Alpha was .79 for the pre-test and .84 for the post-test.

Shame was measured before and after the tasks with six shame items from referring scales of the German versions of the Differential Emotion Scale (DES) [29], [30] and the Positive And Negative Affect Schedule (PANAS) [31], [32]. Two own items ("indignant" and "abashed") were added. To avoid priming, especially before the tasks, we included 34 other items of the DES as well as the PANAS. Items had to be answered on a scale ranging from 1 (*not at all*) to 5 (*extremely*). Cronbach's Alpha was .90 for the pre-test and .93 for the post-test.

General Trust was measured on a 5-point scale with six items [33], e.g., "Most people are trustworthy.". Cronbach's Alpha was .66.

Personality was measured with the German version of the Big Five Inventory (BFI) [34]. For the self-assessment of openness, conscientiousness, neuroticism, agreeableness, and extraversion, 42 items had to be rated on a 5-point scale ranging from 1 (*disagree strongly*) to 5 (*agree strongly*). Cronbach's Alpha for the five scales was between .78 and .90.

IV. Results

In general, 45% of the subjects fulfilled all 18 tasks. The most frequent breaking task was throwing down a coffee cup (17%, task 15), followed by telling a joke (15%, task 6) and taking out a booger from the nose (13%, task 17). On average, 14.35 (*SD* = 5.00) tasks were performed.

A. Assessment of the instructor

The human instructor (M_{Att} = 3.67, SD_{Att} = 0.88; M_{Sym} = 3.50, SD_{Sym} = 0.90) was assessed more attractive ($F(1,56) = 21.56$, $p < .001$, $\eta_p^2 = .278$) and more sympathetic ($F(1,56) = 5.17$, $p < .05$, $\eta_p^2 = .085$) than the virtual agent (M_{Att} = 2.63, SD_{Att} = 0.85; M_{Sym} = 2.93, SD_{Sym} = 1.05). Regarding their perceived authority the human instructor did not differ ($F(1,56) = 0.36$, $p = .48$) from the virtual agent.

Moreover, the trust in the instructor did not differ between the groups ($F(1,58) = 2.90$, $p = .09$). However, trust in the instructor correlated significantly with the perceived shame ($r = .32$, $p < .001$) and stress ($r = .30$, $p < .05$) of the participants.

B. Hypotheses

Hypotheses 1 and 2b tested for a non-existent difference. Therefore, the classic statistical tests, a t-test and a multivariate analysis of variance, is enriched with the Bayes Factor, that allows researchers to express preference for either the null hypothesis or the alternative [35].

Hypothesis 1 stated that obedience, measured with the breaking task, does not differ between the human and agent instructor. We found no significant difference for the breaking task between the conditions ($t(58) = -0.13$, $p = .90$). The Bayes Factor was in favor for the null hypothesis (*JSZ-B_{01}* = 5.11, *Scaled-Information-B_{01}* = 3.97). With the virtual agent, participants finished on average 14.27 (*SD* = 4.90) tasks and with the human instructor 14.43 (*SD* = 5.18). Hence, hypothesis 1 was supported by our data.

Hypothesis 2a proposed that obedience to stressful and shameful tasks leads to a higher self-reported stress and shame level. Our data (Table II for descriptive data) showed that after the experiment the self-reported stress and shame values are significantly higher than before ($F_{\text{Stress}}(1,59) = 12.33$, $p < .001$, $\eta_p^2 = .173$; $F_{\text{Shame}}(1,59) = 60.49$, $p < .001$, $\eta_p^2 = .506$). Therefore, hypothesis 2a was supported by our data.

Hypothesis 2b posited that the level of stress and shame after the experiment does not differ between the groups. Adding condition in the multivariate model used for 2a, we could not find a significant difference between the agent and the

human instructor (*Wilks-*λ = .978, $F(2,57) = 0.63$, $p = .534$). Neither the single t-tests, needed for the Bayes Factor, did show a significant difference between the conditions for stress ($t(58) = -0.70$, $p = .487$) or shame ($t(58) = -0.34$, $p = .739$). The Bayes Factor showed a preference for the null hypothesis for stress (*JSZ-*B_{01} = 4.12, *Scaled-Information-*B_{01} = 3.17) as well as for shame (*JSZ-*B_{01} = 4.89, *Scaled-Information-*B_{01} = 3.79). Overall, hypothesis 2b that the agent invoked the same level of stress and shame like the human instructor was supported by our data.

Table II
DESCRIPTIVES FOR SELF-REPORTED SHAME AND STRESS BEFORE AND AFTER THE TASK FULFILLMENT.

	Agent *M* (*SD*)	Human *M* (*SD*)	Overall *M* (*SD*)
$\text{Stress}_{\text{pre}}$	1.86 (0.47)	2.03 (0.53)	1.94 (0.50)
$\text{Stress}_{\text{post}}$	2.23 (0.69)	2.23 (0.58)	2.23 (0.63)
$\text{Shame}_{\text{pre}}$	1.43 (0.62)	1.61 (0.69)	1.52 (0.66)
$\text{Shame}_{\text{post}}$	2.47 (1.11)	2.42 (0.96)	2.44 (1.03)

Note. $N = 60$. Stress was measured on a 4-point scale, shame was measured on a 5-point scale. Pre stands for the self-reported values before the task fulfillment, post for the self-reported values after the task fulfillment.

Hypothesis 3 stated that there is an interdependency between obedience and personality. The correlations of the five personality factors with obedience, measured with the breaking task, did not reach significant levels. The hierarchical linear regression with the five factors ordered descending according to the strength of the correlation [36], did not reach significance.

V. DISCUSSION

This study aimed to compare a virtual agent giving instructions to fulfill stressful and shameful tasks with a human instructor. Both instructors were presented in a video-chat under the cover-story of a creativity test. Participants had to obey to a maximum amount of 18 tasks increasing in difficulty. The difficulty level was empirically justified in a pre-study. Our results show that participants obey the virtual agent at the same level as the human instructor. On average, around 14 tasks were fulfilled in both conditions. Moreover, we found that obedience in fulfilling shameful tasks increases the level of stress and shame. This increase was independent from the instructor (human vs. virtual agent). The virtual agent and the human invoke feelings of stress and shame to the same amount. Additionally, we examined the influence of personality on obedience to shameful tasks, but could not find any effects.

Our finding that participants obey towards non-humans like towards human instructors is consistent with previous work examining robots as instructors [3], [5], [6].

Likewise the classic and more recently adapted obedience experiments [13], [17], we found that obedience in fulfilling shameful tasks influences participants' self-reported stress and shame. The level of stress and shame increased significantly from before to after the task fulfillment.

Several studies conclude that with a higher degree of anthropomorphism people feel shame also towards non-human entities like virtual characters or robots [18]–[20]. In our study, we could find similar results. Participants reported shame in the post-questionnaire when instructed by a virtual agent, although shame is an interpersonal emotion, i.e., it occurs typically or almost only in the presence of an emotional relation to another human [37]. Even more, the virtual agent does not only invoke shame, but it also invokes the same amount of shame and stress like the human instructor. This goes in accordance with findings by Menne, who showed that non-human entities are able to invoke the same feelings of shame like humans [3].

Regarding the correlation between personality and obedience to shameful tasks, our data did not show significant correlations of the five factors openness, conscientiousness, neuroticism, agreeableness, and extraversion with the amount of fulfilled tasks. The hierarchical linear regression did not show any significant predictions when ordering the factors descending according to the strength of the correlation. Those findings are in line with previous studies [14], [21], but not with [22], where a positive correlation between agreeableness and obedience was found.

A. Limitations and Future Work

Likewise the virtual agent, the human instructor was presented in a video-chat setting. Compared to other studies that had real humans in face-to-face interactions, a human in a video-chat might not have the same authority. A simple reason for this might be that it is not physically present and therefore a smaller threat. However, we could show that almost half of the participants fulfilled all tasks even from an entity that is not physically present. Therefore, our study indicates that the physical presence of the instructor might not be that important as Milgram [13] stressed out. Moreover, guaranteeing a standardized behavior of a human instructor seeing 30 people performing, e.g., the chicken dance, is nearly impossible.

For the assessment of stress and shame, we relied on self-reported data of the participants. Our study design did not include objective measures of stress like in [17]. Also, shame was not measured objectively, e.g., by analyzing social signals of shame or shame regulation [38]. Therefore, future work should include the objective analysis of social signals of shame and shame regulation. This observational data can be an appropriate method to validate self-reported shame and stress.

In our experiment, participants had to fulfill shameful tasks. Therefore, we do not know if the increase in the stress and shame levels in both conditions is due to the obedience or due to the execution of the tasks. Future work could consider using more neutral tasks to find out if obedience itself leads to an increase in the stress and shame levels.

This work compares a human and a virtual agent in a video-chat as instructors. Future research could include other instructors like a present robot or a robot in a video chat. Moreover, to examine the influence of anthropomorphism the virtual agent could be compared to a conversational agent without representation.

VI. Conclusion

The results of this study indicate that humans obey virtual agents just as they do towards humans in a video-chat. We could show that participants fulfill the same amount of shameful tasks independent of the instructor. Moreover, both instructors were able to elicit the same level of shame and stress in the participant. Therefore, our results provide one more indication for the validity of the Media Equation. Virtual agents seem to be able to influence humans even when it comes to tasks that are uneasy to perform.

Acknowledgment

We thank Charamel GmbH for realizing our requirements with regard to the virtual agent and Saarland Informatics Campus for the provision of the experiment rooms.

References

[1] L. Hansen, "8 drivers who blindly followed their GPS into disaster," *John Guehl*, 2013. [Online]. Available: https://theweek.com/articles/464674/8-drivers-who-blindly-followed-gps-into-disaster

[2] A. Lunney, N. R. Cunningham, and M. S. Eastin, "Wearable fitness technology: A structural investigation into acceptance and perceived fitness outcomes," *Computers in Human Behavior*, vol. 65, pp. 114–120, 2016.

[3] I. M. Menne, "Yes, of course? an investigation on obedience and feelings of shame towards a robot," in *International Conference on Social Robotics*. Springer, 2017, pp. 365–374.

[4] M. C. Gombolay, R. A. Gutierrez, S. G. Clarke, G. F. Sturla, and J. A. Shah, "Decision-making authority, team efficiency and human worker satisfaction in mixed human–robot teams," *Autonomous Robots*, vol. 39, no. 3, pp. 293–312, 2015.

[5] D. Y. Geiskkovitch, D. Cormier, S. H. Seo, and J. E. Young, "Please continue, we need more data: an exploration of obedience to robots," *Journal of Human-Robot Interaction*, vol. 5, no. 1, pp. 82–99, 2016.

[6] D. Cormier, G. Newman, M. Nakane, J. E. Young, and S. Durocher, "Would you do as a robot commands? an obedience study for human-robot interaction," in *International Conference on Human-Agent Interaction*, 2013.

[7] B. Reeves and C. I. Nass, *The media equation: How people treat computers, television, and new media like real people and places.* Cambridge University Press, 1996.

[8] I. M. Menne and B. Lugrin, "In the face of emotion: a behavioral study on emotions towards a robot using the facial action coding system," in *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2017, pp. 205–206.

[9] J. E. Young, J. Sung, A. Voida, E. Sharlin, T. Igarashi, H. I. Christensen, and R. E. Grinter, "Evaluating human-robot interaction," *International Journal of Social Robotics*, vol. 3, no. 1, pp. 53–67, 2011.

[10] T. Schneeberger, P. Gebhard, T. Baur, and E. André, "Parley: a transparent virtual social agent training interface," in *Proceedings of the 24th International Conference on Intelligent User Interfaces: Companion*. ACM, 2019, pp. 35–36.

[11] D. DeVault, R. Artstein, G. Benn, T. Dey, E. Fast, A. Gainer, K. Georgila, J. Gratch, A. Hartholt, M. Lhommet *et al.*, "Simsensei kiosk: A virtual human interviewer for healthcare decision support," in *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-Agent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 2014, pp. 1061–1068.

[12] R. Niewiadomski, E. Bevacqua, M. Mancini, and C. Pelachaud, "Greta: an interactive expressive eca system," in *Proceedings of the 8th International Conference on Autonomous Agents and Multi-Agent, Systems-Volume 2*. International Foundation for Autonomous Agents and Multiagent Systems, 2009, pp. 1399–1400.

[13] S. Milgram, "Some conditions of obedience and disobedience to authority," *Human Relations*, vol. 18, no. 1, pp. 57–76, 1965.

[14] J. M. Burger, "Replicating milgram: Would people still obey today?" *American Psychologist*, vol. 64, no. 1, p. 1, 2009.

[15] S. Agrawal and M.-A. Williams, "Robot authority and human obedience: A study of human behaviour using a robot security guard," in *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2017, pp. 57–58.

[16] J. K. Burgoon, J. A. Bonito, B. Bengtsson, C. Cederberg, M. Lundeberg, and L. Allspach, "Interactivity in human–computer interaction: A study of credibility, understanding, and influence," *Computers in Human Behavior*, vol. 16, no. 6, pp. 553–574, 2000.

[17] M. Slater, A. Antley, A. Davison, D. Swapp, C. Guger, C. Barker, N. Pistrang, and M. V. Sanchez-Vives, "A virtual reprise of the stanley milgram obedience experiments," *PloS one*, vol. 1, no. 1, p. e39, 2006.

[18] C. Bartneck, T. Bleeker, J. Bun, P. Fens, and L. Riet, "The influence of robot anthropomorphism on the feelings of embarrassment when interacting with robots," *Paladyn, Journal of Behavioral Robotics*, vol. 1, no. 2, pp. 109–115, 2010.

[19] G. M. Lucas, J. Gratch, A. King, and L.-P. Morency, "Its only a computer: Virtual humans increase willingness to disclose," *Computers in Human Behavior*, vol. 37, pp. 94–100, 2014.

[20] S.-H. Kang and J. Gratch, "Virtual humans elicit socially anxious interactants' verbal self-disclosure," *Computer Animation and Virtual Worlds*, vol. 21, no. 3-4, pp. 473–482, 2010.

[21] S. Milgram, *Obedience to Authority: An Experimental View*. Harper & Row, 1974.

[22] L. Bègue, J.-L. Beauvois, D. Courbet, D. Oberlé, J. Lepage, and A. A. Duke, "Personality predicts obedience in a m ilgram paradigm," *Journal of Personality*, vol. 83, no. 3, pp. 299–306, 2015.

[23] J. A. Abe, "Shame, guilt, and personality judgment," *Journal of Research in Personality*, vol. 38, no. 2, pp. 85–104, 2004.

[24] D. W. Harder and D. F. Greenwald, "Further validation of the shame and guilt scales of the harder personal feelings questionnaire-2," *Psychological Reports*, vol. 85, no. 1, pp. 271–281, 1999.

[25] D. Einstein and K. Lanning, "Shame, guilt, ego development and the five-factor model of personality," *Journal of Personality*, vol. 66, no. 4, pp. 555–582, 1998.

[26] P. Gebhard, T. Baur, I. Damian, G. Mehlmann, J. Wagner, and E. André, "Exploring interaction strategies for virtual characters to induce stress in simulated job interviews," in *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-Agent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 2014, pp. 661–668.

[27] P. Gebhard, G. Mehlmann, and M. Kipp, "Visual scenemakera tool for authoring interactive virtual characters," *Journal on Multimodal User Interfaces*, vol. 6, no. 1-2, pp. 3–11, 2012.

[28] T. M. Marteau and H. Bekker, "The development of a six-item short-form of the state scale of the spielberger statetrait anxiety inventory (stai)," *British Journal of Clinical Psychology*, vol. 31, no. 3, pp. 301–306, 1992.

[29] C. E. Izard, *The Differential Emotions Scale: DES IV-A;[a Method of Measuring the Meaning of Subjective Experience of Discrete Emotions]*. University of Delaware, 1993.

[30] J. Merten and R. Krause, *DAS (Differentielle Affekt Skala)*. Univ., Fachbereich Sozial-und Umweltwiss., Fachrichtung Psychologie, 1993.

[31] D. Watson, L. A. Clark, and A. Tellegen, "Development and validation of brief measures of positive and negative affect: the panas scales." *Journal of Personality and Social Psychology*, vol. 54, no. 6, p. 1063, 1988.

[32] H. W. Krohne, B. Egloff, C.-W. Kohlmann, and A. Tausch, "Untersuchungen mit einer deutschen version der" positive and negative affect schedule"(panas)," *Diagnostica-Gottingen-*, vol. 42, pp. 139–156, 1996.

[33] R. M. Kramer, "Trust and distrust in organizations: Emerging perspectives, enduring questions," *Annual Review of Psychology*, vol. 50, no. 1, pp. 569–598, 1999.

[34] O. P. John, E. M. Donahue, and R. L. Kentle, "The big five inventoryversions 4a and 54," 1991.

[35] J. N. Rouder, P. L. Speckman, D. Sun, R. D. Morey, and G. Iverson, "Bayesian t tests for accepting and rejecting the null hypothesis," *Psychonomic Bulletin & Review*, vol. 16, no. 2, pp. 225–237, 2009.

[36] A. Field, *Discovering statistics using IBM SPSS statistics*. Sage, 2013.

[37] C. E. Izard, *Human emotions*. Springer Science & Business Media, 2013.

[38] P. Gebhard, T. Schneeberger, T. Baur, and E. André, "Marssi: Model of appraisal, regulation, and social signal interpretation," in *Proceedings of the 17th International Conference on Autonomous Agents and Multi-Agent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 2018, pp. 497–506.