

A Bizarre Virtual Trainer Outperforms a Human Trainer in Foreign Language Word Learning

Manuela Macedonia^{1,2}

¹ Department for Information Systems, University of Linz, Austria

² Research Group Neural Mechanisms of Human Communication,
Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany
manuela@macedonia.at

Abstract-In this study, the effects that a human trainer and a pedagogical virtual agent have on the memory for words in a foreign language (L2) were investigated. In a recent study on L2 word learning, Bergmann and Macedonia (2013) cued participants to memorize novel words both audiovisually and by performing additional gestures. The gestures were performed by both a human and a virtual trainer. In some of the tests, the virtual agent had a greater positive influence on memory performance than the human trainer. In order to determine why the agent was a better trainer than the human, 18 naive subjects were invited to rate the gestures performed by both trainers. Furthermore, participants were asked to evaluate their perception of the human and the agent. It was hypothesized that the gestures performed by the agent would be more peculiar than those by the human and possibly attract greater attention. It was also hypothesized that the agent's personality might be more appealing than that of the human. The results showed that the agent's gestures were perceived as less natural than those of the human. This might have triggered greater attention and/or more emotional involvement of the participants. The perception of both trainers as "personalities" did not differ, with the exception of a few traits for which the human trainer was considered to be better. Altogether, because of the peculiar gestures it made and because of its looks, the agent may have been perceived as bizarre. Therefore, he might have induced the bizarreness effect in the memory for words.

Keywords- *Virtual Agent; Evaluation; Foreign Language; Learning*

I. INTRODUCTION

A. Pedagogical Agents and Their Impact on Learning

In the last decade, great efforts have been put into the development of pedagogical agents, i.e. humanlike, virtual characters that are employed in educational settings to support and facilitate learning [1]. The idea that we have pedagogical agents is that their presence enhances learners' motivation, which in turn might lead to more effective learning [2, 3]. Pedagogical agents look like humans and in a certain way they are also capable of some kind of social interaction, e.g., by nodding their heads, looking at the user and producing facial expressions [4]. With their gestures they can also direct attention to contents that have to be learned and guide users through them [5]. Motivational words of the agents support users in their learning tasks [6]. The efforts put into making the agents resemble humans and enabling them to perform interactive behavior have been made in order to attain what is known as the *persona effect* [7], i.e. the positive effect that a lifelike character has on learners [8]. It is known that a virtual character with anthropomorphic looks is likely to be accepted by learners [9]. However, findings on the benefit of an agent on the learning outcome of humans have been inconclusive [1]. In some studies, pedagogical agents improve learning and in other studies they do not. Positive effects were shown, for example, in an experiment by Beun and colleagues [10], where participants were provided with a short story, presented either by (1) a realistic virtual human, by (2) a cartoon-like virtual character, or (3) as a text. Both agent conditions yielded better results than the one without an agent with respect to short-term memory performance. Similarly, a recent study demonstrated that learners performed better on a transfer test when a human-voiced agent displayed rich nonverbal behavior (human-like gestures, facial expression, eye gaze, and body movement) than when it did not do so [11].

On the other hand, in a small number of studies, no beneficial effects of pedagogical agents on the learning outcome were seen. For instance, in an experiment by Mondriou and Virvou [12], participants were assisted by an anthropomorphic agent that guided them and provided them with supportive feedback in solving algebraic problems. There, the presence of the virtual character did not improve short-term learning results. Similarly in another study, Mulken et al. [13] utilized a virtual agent which pointed a stick at verbal information that the subjects were supposed to retrieve from texts on machines and texts on fictitious persons, i.e. they compared the effects of performing deictic gestures with a no-agent condition. Recall tests showed that the agent's gestures had no impact on memory retrieval. To our knowledge, the only vocabulary learning experiment conducted with an anthropomorphic agent as a trainer is the between-subject study by Miksatko et al. [14], in which 36 participants learned English vocabulary terms. The modality of training was a flash card vocabulary trainer imparting vocabulary items like animals, clothes, body and medicine. It displayed a written word in German and its correspondent in English framed in a silver screen. In the agent-condition, additionally, a female agent stood between the two cards, moved slightly, pointed to the cards and gave verbal positive feed-back by nodding or smiling. In the no-agent condition, the machine gave written positive feedback. Participants learned equally well under both conditions. Thus, the authors of the study could

not find any memory enhancement due to the presence and feedback of the agent and coined the expression “persona-zero effect”, which means “that adding an agent does not benefit the performance but also, does not reduce it” [14].

In all studies considered here so far, two main aspects of learning - the method and the trainer - were not controlled. In other words, there was interaction between them. On the one hand, we find the manner in which contents are presented and trained, for example the kind of instructions (inductive or deductive), the duration of the exposure to the stimuli, the number of repetitions, the visual input, and so on. On the other hand, we have the user-agent interaction that also should have an impact on learning. The way the agent says something, the time when it says it and similar communicative features certainly influence learning outcomes, as was demonstrated by Kim et al. [15]. Thus, both aspects are intertwined in experiments with multimodal agent interaction, and training results cannot be neatly attributed to either the kind of training or the human-agent social interaction. In fact, this is more or less what happens in natural classroom interaction; the role of the teacher is one factor in the learning processes that decisively influences the results [16, 17]. Independent of the different procedures used in encoding, instructors interact socially with their learners, and hence influence their learning behavior. In learning situations, the encoding and the interaction between learners and trainers are closely interconnected. Under normal circumstances, one cannot clearly attribute learning results to either encoding or social interaction. At the same time, it is not possible to discern the extent to which encoding and interaction influence each other.

B. The Method: Enactment

In a previous study of the authors [18], we opted for enactment, a method that is not largely known and practiced in which learners perform a gesture related to the word they have to acquire in foreign language (L2) [19]. In the study, the trainers (agent and human) cued participants to produce an iconic gesture when learning novel words in L2. In L2 instruction, learners usually read, listen, write and (re)produce the language to be learned [20]. In the case of vocabulary, items are introduced in dialogues and texts, elaborated in exercises and various classroom activities and traditionally learned from lists. Note that since the early eighties, laboratory research has shown that accompanying words or phrases with a gesture enhances their memorization [21]. The effect that gestures have on the retention of verbal information was termed the enactment effect [22] and has been demonstrated not only for words but also for phrases and sentences, as well as in different populations. Enactment has also been documented for foreign language (L2) and has proven to have a robust effect, especially in the long term [19]. At least four explanations account for this. Firstly, enacting a word leaves a motor trace in one's memory [23]. This has been convincingly proven in neuroscientific experiments in which it is possible to detect activity in motor areas upon acoustic or visual word presentation [24, 25]. Secondly, enhancement results from motor imagery [26, 27]: it leads to deeper information processing [28] and has positive effects on memory. Thirdly, gesture performance during word learning enhances the attention devoted to that word. More attention is paid to it than when it is learned from a list [29] and this factor has an impact on memory. Fourthly, by learning verbal information through enactment the representation of the word becomes more complex [30-32]. This again leads to deeper processing [28] than only reading or reading and listening to the words does.

C. The Trainer and His Impact on Memory

If enactment is an effective way of training vocabulary and if it can successfully replace tedious words lists, enactment should be accessible to learners on a large scale and in flexible training settings. Therefore, in our initial study [18], we evaluated the effect of using a pedagogical virtual agent as a vocabulary trainer. It is a computer interface with an anthropomorphic appearance that enunciates the items to be learned in L2 and performs semantically related gestures in a way similar to that of a human trainer. We reasoned that an agent could replace a human. In fact, he could provoke the persona-effect [7] and thereby bring about an enhancement in the same way a human trainer does. Furthermore, unlike a living teacher, a virtual agent is at the user's disposal whenever he wants and needs to train vocabulary items. It “resides” in the user's computer, tablet or mobile phone and can support training at any time of the day, 365 days a year. However, no tests have been done to determine whether the agent would produce the same effects as a human trainer; effects of a human trainer had already been well documented in other experiments [32, 33].

Briefly, in a within-subject study, 29 subjects learned 45 novel words in Vimmi, an artificial corpus created for experimental purposes [18]. The words conformed to Italian phonotactics and were assigned common meanings such as bridge and necklace. They were equally distributed among three training conditions: a baseline in which participants only heard the words and read them aloud and two conditions in which participants were additionally cued to imitate the gestures performed by either the virtual agent or a human actor. The iconic gestures mirrored some feature of each word's semantics that were chosen by the experimenter. The training lasted three days (1-3). Memory performance was measured on days 2-4 and on day 30 by means of free and cued recall tests. In the short term, the repeated-measures ANOVA with the within-subject factors TIME (day 01-04) and TRAINING (human trainer and agent trainer) showed a significant main effect of TIME on the free recall $F(2,56)=187.26$, $p<.001$. Words trained with the agent were retrieved significantly better than ones that were trained in the baseline-condition $F(1, 28) = 6.25$, $p=.019$. There was however no significant difference between words learned with the agent and those learned with the human trainer. In the long term, because of the high variability among learners, the population test group was subdivided into high and low performers by means of a median-split. A significant main effect of TRAINING $F(2, 54) = 3.68$, $p=.032$ was found in the free recall. Items were retrieved better when they were learned with the virtual agent $F(1, 27) = 9.17$, $p=.005$ than with the human trainer $F(1, 27) = 4.72$, $p=.039$. The data also revealed a significant interaction

effect between TRAINING and PERFORMANCE, $F(2, 54) = 4.37$, $p = .017$, and indicated that the different training conditions had different impacts on low and high performers. In the cued recall, the primary effect of TRAINING was significant, $F(2, 54) = 3.49$, $p = .038$. Words trained with the agent were retrieved significantly better ($F(1, 27) = 5.90$, $p = .022$) than the ones trained in the baseline situation. Again an interaction between TRAINING and PERFORMANCE was found, $F(2, 54) = 3.99$, $p = .024$, in high performers.

Overall, this experiment confirmed the results of previous studies on the enhancing effect of gestures on memory of verbal information. Furthermore, the data showed that the memory outcomes of training with the agent did not differ from those with the human and in some cases the impact of the virtual agent on performance was greater than that of the human. In order to investigate possible trainer effects of the agent, data of the perception of the gestures performed by the agent and by the human and of the perception of both trainer “personalities” were obtained. It was hypothesized that participants had perceived a) the agent’s gestures as peculiar and b) the agent’s personality as pleasant. Thus, the aim of the present study was to evaluate the effect that the trainer’s gestures and personality had on the participant’s performance.

II. METHOD

A. Participants

Eighteen participants (11 male, 7 female), with a mean age of 30.38 years (SD 11.85), took part in the study. They were recruited at Bielefeld University and Linz University. Participants were asked to participate for free but had a chance of winning 5 gift coupons, each with a value of 10€, for an online-bookshop. The participants were naive of the study’s goal.

B. Stimuli

In Bergmann and Macedonia [18], participants were trained by 30 videos. Each of the videos was 4 s in length. In 15 of the videos, a human actress performed the gestures. In the other 15 videos, Billie, the agent, performed them. The agent’s gestures were specified in the Multimodal Utterance Representation Markup Language (MURML) [34] and realized by the Articulated Communicator Engine (ACE) [35], a toolkit for making animated embodied agents that are capable of generating human-like multimodal utterances. The agent’s gestures were rendered into videos (Fig. 1 a and 1 b). All of the gestures were randomly presented during the online-survey in two blocks, one for the human and one for the agent.

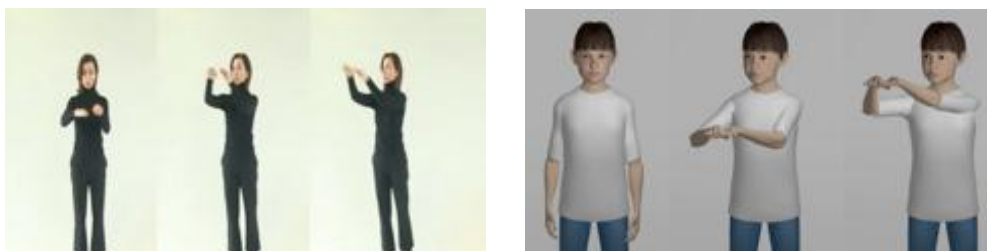


Fig. 1 a and 1 b. Screenshots of the videos with the human and the agent’s gestures for the Vimmi word *gelori* (staircase)

C. Sampling Procedure

Participants were asked to complete the online survey. They were instructed to carefully watch every single video presented and rate it according to the questions it contained. The survey was introduced by a short description in which participants had to provide a few personal details (age, place of residence, school or university degree). Thereafter, the questionnaire was divided into two subsections. In the first subsection, information about the quality of the gestures performed by the agent and by the human was collected; in this regard, we used the attributes iconicity of the gesture, naturalness of the gesture, speed of gesture production and fluidity of gesture production as previously done in another study by Bergmann et al [36]. In the second subsection, participants were asked to rate their perception of both trainers’ personalities in respect to the following characteristics: pleasantness, friendliness, reliability, intelligence, competence, human-likeness and sympathy [37, 38]. At the end of the questionnaire, participants were asked if they would consider working with the human and the agent trainer as L2 teachers. This question could be only answered with yes or no.

III. RESULTS

In the following sections, results regarding the quality of the gestures and the perception of the trainer’s personality are presented. Variables were rated on a five-point Likert scale, with 1 standing for the best evaluation and 5 for the worst one.

A. Quality of Gestures

In order to test the effect of the quality of the gestures, we conducted a repeated measure, 4x2 ANOVA with the four factors affecting the QUALITY OF GESTURES (iconicity, naturalness, speed of execution, fluidity of execution) and the factor TRAINER (human and agent). Significant main effects were found for both experimental factors, i.e. QUALITY OF

GESTURES ($F(3, 51)=35,62$, $p<0.001$) and TRAINER ($F(1,17)=86,6$, $p<0.001$), and a significant interaction was also observed between gestures and trainer ($F(3, 51)=22,01$, $p<0.001$), see Fig. 2.

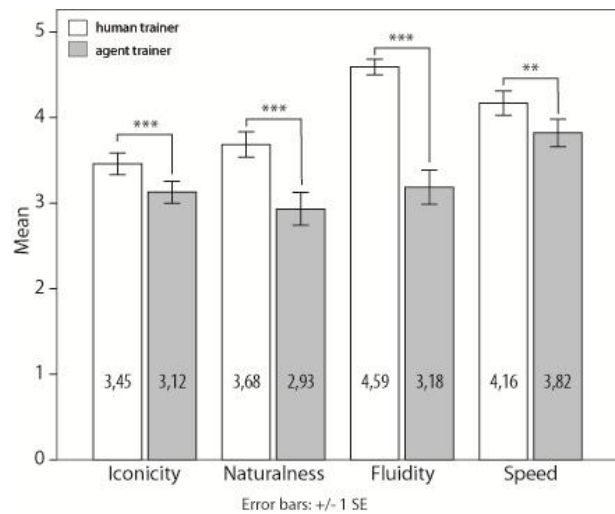


Fig. 2 Rating of the trainers' gestures

1) Iconicity of Gesture

This factor quantifies the extent to which the gestures represent the words' semantics. Agent's gestures were modelled on the base of videos showing the human actress performing them. Thus, no significant difference was expected. However, on the 5-point Likert-scale (too few-too many) the overall mean value for the iconicity of gesture was $M=3.45$ ($SD=.53$) for the human and $M=3.12$ ($SD=.54$) for the agent. Furthermore, for the dependent variable iconicity of gesture, the ANOVA yielded significant results. The iconicity of the gestures the human produced was perceived as significantly better ($F(1, 17) = 42.28$, $p<0.001$).

2) Naturalness of Gesture

With regard to the evaluation of the naturalness of the gestures, the participants rated as follows: human $M=3.68$ ($SD=.63$), agent $M=2.93$ ($SD=.81$). A significant difference in the factor naturalness was found between the human's and the agent's gestures, $F(1, 17) = 49,34$, $p<0.001$. In fact, even if the agent resembles a human, his gestures still look "robotic". Therefore, they must have been perceived different by the raters (see supplementary video materials online).

3) Fluidity of Gesture Production

For the dependent variable fluidity of gesture production, the participants rated the human gestures as more fluid than those of the agent, with $M=4.59$ ($SD=.38$) and $M=3.18$ ($SD=.85$), respectively. Also, this factor was significant ($F(1, 17) = 58,28$, $p<0.001$).

4) Speed of Gesture Production

The analysis of the ratings for the speed of gesture production showed that the human had a better performance, $M=4.16$ ($SD=.61$), than the agent, $M=3.82$ ($SD=.68$). The ANOVA again showed a significant effect of this factor ($F(1, 17) = 12,56$, $p=.002$).

B. Perception of the Trainer's Personality

With regard to the perceived factor PERSONALITY of the trainer (Figure 3), a repeated measure 7×2 -ANOVA was conducted, i.e. pleasantness, friendliness, reliability, intelligence, competence, human resemblance, and sympathy and TRAINER (human and agent). In contrast to what was predicted, no significant primary effects of the factor PERSONALITY were found ($F(6, 102) = 720$, $p<.634$). However, the factor TRAINER was significant ($F(1, 17) = 10,4$, $p<0.005$) and there was also a significant interaction between the PERSONALITY and the TRAINER ($F(6, 102) = 6,92$, $p<0.001$).

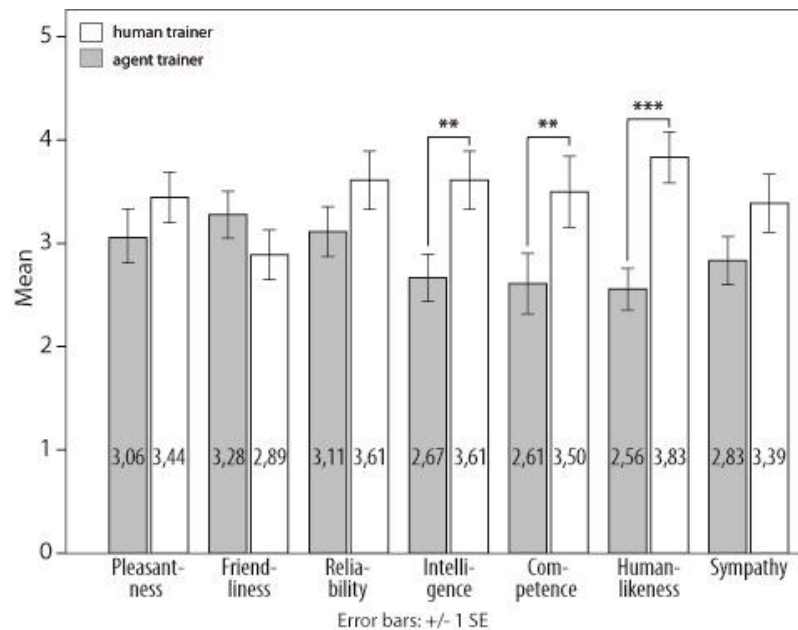


Fig. 3 Rating of the trainers' personality perception

1) Pleasantness

The participants were asked to rate the pleasantness of the agent and the human as an indicator of their attitude towards the trainer. Participants found the human trainer more pleasant, $M=3.44$ ($SD=1.04$), than the agent, $M=3.06$ ($SD=1.11$). However, the difference was not significant.

2) Friendliness

With regard to perceived friendliness, the agent scored better, $M=3.28$ ($SD=.95$), than the human, $M=2.89$ ($SD=1.02$), but this result only reached chance level.

3) Reliability

Another variable of interest, also important in a learner-teacher interaction, is reliability. Of course, if one is to trust him / her, one must assume that what he / she teaches is correct. The agent was considered less reliable, $M=3.11$ ($SD=1.02$), than the human trainer, $M=3.61$ ($SD=1.19$), however not significantly so.

4) Intelligence

Regarding intelligence, the human trainer made a more intelligent impression on the raters, $M=3.61$ ($SD=1.19$), than the agent, $M=2.67$ ($SD=.97$). For this character trait, a significant main effect ($F(1, 17) = 18.26$ $p < 0.01$) was observed.

5) Competence

For the dependant variable competence, raters agreed that the human trainer was more competent, $M=3.50$ ($SD=1.46$), than the agent, $M=2.61$ ($SD=1.46$). The difference between the mean evaluation of the competence of the trainers was significant ($F(1, 17) = 9.37$ $p < 0.01$).

6) Human-likeness

The participants were also asked to rate the degree of resemblance to a human. As one might expect, the human trainer scored much better, $M=3.83$ ($SD=1.04$), than the agent, $M=2.56$ ($SD=0.85$), in this trait. The analysis of the ratings for human-likeness revealed a main effect ($F(1, 17) = 28.36$ $p < 0.001$).

7) Sympathy

Finally, the raters had more sympathy with the human trainer $M=3.83$ ($SD=1.04$), than with the agent, $M=2.83$ ($SD=1.04$). However, for the factor sympathy, the difference was not significant.

C. Human or Agent Trainer?

At the end of the online-survey, a final question asked the participants to indicate their preference for either the human or the agent trainer on the Likert scale. The human trainer scored better, $M=3.56$ ($SD=1.09$), than the agent, $M=2.39$ ($SD=1.19$), and this difference was significant ($F(1, 17) = 20.31$ $p < 0.001$), see Fig. 4.

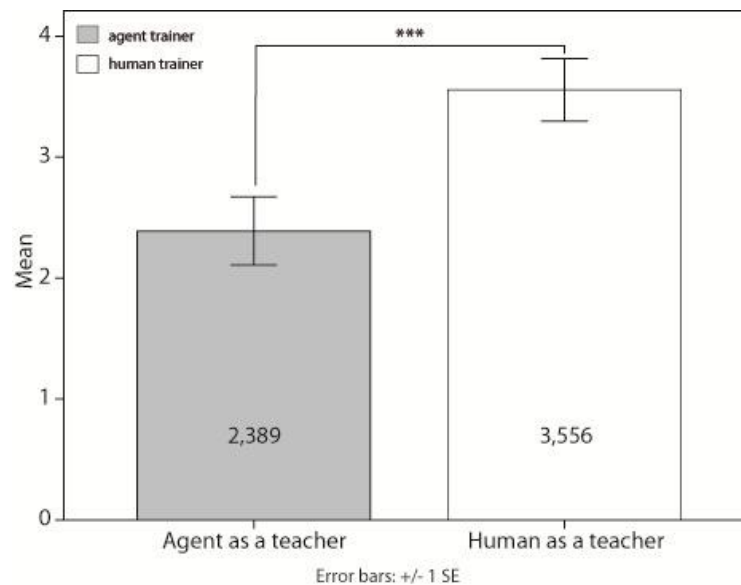


Fig. 4 Rating of the trustworthiness of the teacher

IV. DISCUSSION

The present study was designed to investigate the impact of an agent on word learning. Specifically, we were interested in the quality of the gestures and the perception of the trainer - human or agent - influence learning performance. Participants rated the quality of human gestures significantly higher than the quality of the agent's gestures. The overall perception of both trainers' personalities did not significantly differ, although the human scored better in a few traits. Interestingly, the agent was perceived as more friendly. Note that neither the agent nor the human gave any feedback nor interacted in any way with the raters (i.e. the learners). Thus, this preference may simply have to do with the looks of the trainers and the way they enunciated the words to be learned. Also, the participants answered that they would prefer to be trained by the human teacher than by the agent. Altogether, these results can be interpreted to mean that participants would feel more comfortable with the human than with the agent. Consequently, one would assume that participants should have learned better with the human trainer. However, this was not the case. As reported in the introduction, in free recall tests, high performers learned significantly better with the agent than with the human trainer [18]. Hence, the results reported in the present study seem to be paradoxal. In the following sections, the reasons that might explain why high performers had better memory results when trained by the agent will be discussed.

A. How Are Non-Human Gestures Processed in the Brain?

A small number of neuroscientific studies have investigated the brain's responses upon presentation of robotic movements by means of functional magnetic resonance imaging (fMRI) or electroencephalography (EEG). fMRI records changes in oxygenated blood flow in the brain and indicates the topography of cognitive processes that take place during stimulation. It is well known that watching gestures elicits activity in motor areas of the brain and in mirror neuron circuits [39]. In fact, mirror neurons, a particular class of neurons in the parietal and premotor regions of the brain, become active when a person executes an action and also when the person observes the execution of the same action by others. In a study by Gazzola et al. [40], participants watched actions performed by both a human and an industrial robot while lying in the fMRI-scanner. Both the human and the robot elicited brain responses in resonating mirror neuron circuits, which in both cases indicated that they had understood the action. Most importantly, the brain responses did not differ from each other, i.e. the brain interpreted gestures performed by the robot as if they had been performed by the human, even though the robot's appearance was not human-like.

EEG, another method used in neuroscience, detects electrical activity on the scalp and thereby elucidates the temporal dynamics of cognitive processes. In order to determine whether the brain is involved in the processing of sensorimotor experience, so called neural oscillations are recorded, i.e. rhythmic activity patterns generated by neurons. If sensorimotor processing occurs, there is a sudden pattern changes in the motor cortex cells: the so called mu-rhythm - a brain wave present when the body is at rest - is suppressed. The suppression of the mu-rhythm is detected when a person performs a motor act but also sees somebody else doing it. Oberman et al. [41] recorded oscillations of brains upon presentation of robotic actions through videos. The authors were able to detect the suppression of mu-rhythms, indicating that motor regions of the brain had become active. It seems that, if robotic movements are modelled according to human biological motion, human perception is not greatly affected [42].

Differences in perception do occur when the movements of virtual figures strongly diverge from biological motions. In an fMRI-study by Chaminade et al. [43], the responsiveness to robotic movements decreased in resonating motor networks along

with the reduction of biological features. However, it is interesting to note that in this experiment, areas in the forebrain usually engaged in mentalizing became active. Here, subjects might have unconsciously tried to match the motions perceived with a stored (human) motion template and to make a sense out of them.

Moreover, studies with electromyography - a physiological method used to record electrical activity in muscles - have demonstrated that human body responds to the same extent no matter a motion is produced by a human or a robot, as long as the motion produced by the robot is comparable to the human motion [44]. Thus, even if we watch a robotic arm grasping an object with iron pliers, our body is ready to reproduce the movement as long as the brain identifies the pliers as the abstraction of a hand and can match the action of the robot with the experience of grasping [45, 46].

Altogether, neuroscientific studies show that the brain processes human and robotic motions in a similar way and that in doing so it is quite tolerant of divergence from human patterns. It is thus no surprise that the participants of this study were able to learn from the agent's gestures as well as from the human gestures. Even if participants were fully aware that the agent's gestures were not performed as well as those of the human - which they consciously stated in the survey - their learning efficiency was not affected by this discrepancy.

But how can we explain the fact that high performers showed better memory results in some free recall tests for words learned with the agent? If this cannot be explained by the gesture quality, we reason that the factor affecting performance must lie in the perception of the trainer's personality.

B. How Does the Brain Perceive an Agent Trainer in Comparison to a Human Trainer?

When we began the learning experiment [18], the main concern was that the agent would not be as good as the human in training participants because of the fact that it looks like an avatar and because of its robotic gestures (see online supplementary materials). In fact, it has been shown that the human-likeness of robots and avatars has an effect on human perception and preference [47]. That is to say that people interacting with robots prefer human-like appearance and attributes. Besides, it was assumed that humans would learn better with a human trainer on the basis of the results of a study by Chaminade et al. [48]. The authors presented participants with emotional gestures, i.e. facial expressions of anger, joy and disgust performed by a human and by a humanoid robot, and while doing so they measured the participants' neural responses with functional magnet resonance imaging (fMRI). The areas of the brain that processed the emotions mentioned above were reduced when participants saw that a robot was making the facial expressions instead of a human face. We thus assumed that this reduced processing might also occur in this study and affect the learning outcome. A closer look at the literature, however, shows that this study is only one side of the coin. In an experiment that utilized EEG to measure brain responses to human and non-humanoid robots, Dubal et al. [49] showed that emotion evoked by the robot faces is processed as fast as that caused by humans, even if the robots do not have an anthropomorphic appearance. In fact, participants responded equally well to emotional expressions of real and schematic faces [50]. Furthermore, robotic stimuli (even non-humanoid) can elicit stronger attention when they are being processed [51]. Hence, we speculate that our participants processed their personal perception of the agent in a manner very similar to the way they did so for the human trainer. This could explain why the agent was not perceived in a significantly different way than the human trainer. Also, participants in Bergmann and Macedonia's first study [18] - all students of information engineering - might have felt some kind of sympathy with the virtual figure.

A further mechanism might have modulated the outcome of the human-agent interaction, i.e. its distance from the so called uncanny valley [52]. This term was coined by M. Mori, a Japanese robotics professor who described how human acceptance of robots does not increase linearly with the increase in the robots' human-likeness. In fact, when the virtual character becomes progressively more human-like, the acceptance of it does at first increase. Then, at once, the acceptance abruptly slopes down into the uncanny valley, a "cognitive landscape" in which humans perceive discomfort towards the humanlike object and regard it as eerie and strange. In fact, the virtual character does look like a human, but it fails to live up to what one would expect from a real person [53]. When, however, the human-likeness increases to the point that the virtual figure is perceived as real, acceptance of the robot again increases and reaches its maximum. Interestingly, a study by Walters and colleagues [47] also showed that participants who were considered to be introverted and emotionally unstable had preferences for robots that looked more mechanical. In our case, Billie does have an anthropomorphic appearance, but because of different features like motion, tridimensionality, and so on, he is still perceived as a robot. It is possible that Billie's robotic gestures and looks prevented him from coming close to an uncanny valley. This might have biased the acceptance of the agent, who was rated as more sympathetic (although not significantly so) than the human trainer (Figure 3).

In summary, neuroscientific studies have shown that humans perceive human and robotic gestures as well as faces in similar ways. This explains why in Bergmann and Macedonia's study [18], the human and agent trainer had similar effects on participants' memory performance. However, in our survey, participants rated both the humans's gestures and personality as better than those of the agent. That seems to be paradoxal in the light of the fact that Billie overperformed the human trainer when instructing high performers. What can the reason be for that?

C. Is Billie Bizarre?

Considering both factors together without taking the ratings into account: the agent performs gestures in a strange manner

and it is not a human. This might make it particularly salient, or even bizarre. Because of its saliency it might attract more attention, which could enhance memorization. Also its bizarreness might have the same effect. The bizarreness effect on memory refers to the fact that bizarre or unusual information is recalled better than normal information [54]. Two different theoretical positions offer explanations for the bizarreness effect: encoding and retrieval theories. Encoding theories explain the effects on memory on the basis of the fact that bizarreness arouses increased attention and requires deeper processing. Retrieval theories maintain that this effect is a matter of retrieval, and it is very fragile and mainly detected in free recall tests.

First, for encoding, if learners read a common sentence such as “when I went to school I met a lady who was walking her dog”, they will engage fewer cognitive resources than when they read a bizarre sentence like “when I went to school I met a lady who was walking her crocodile”. Similarly, processing a human trainer who is performing gestures and enunciating words in a foreign language might afford less of a challenge than doing the same thing with a virtual trainer that looks and makes gestures like a robot. Hence, it is possible that greater cognitive resources were required to perceive and process Billie as a trainer, and that the robot aroused increased attention. A deeper processing effect [55] of the kind which is found during semantic processing, i.e. the elaboration of verbal information, does not seem to be pertinent here. Both the human and the agent trainer cued participants to simply perform the gestures and to repeat the words they heard.

In order not to be suggestive, we did not expressly ask, whether participants regarded Billie as bizarre. However, this might have been the case. Billie could have induced a bizarreness effect on memory. In fact, it has been shown that bizarre or unusual imagery is retrieved better than common imagery [54, 56]. This is also the case for bizarre orthography [57] and high frequency words [58]. In addition, the bizarreness effect is an explanation for the better recognition of enacted items [59]. It has been reported that this effect is obtained by using a mixed-items design, i.e. by alternately requiring subjects to memorize bizarre and common items [59]. This kind of design was also utilized in our experiment [18]. Participants were cued to train the words under both conditions - human and trainer - during the same training session. As previously mentioned, the bizarreness effect has been investigated for both images and verbal information. In our case, participants learned novel words of Vimmi, an artificial corpus created for experimental purposes. Although the items were all well controlled in respect to their length, the occurrence of unfamiliar phonemes, intonation and syllabic sequences might have sounded peculiar. The fact that the novel words sounded like Italian, might have conveyed some bizarreness into the verbal information. At least, we cannot exclude this possibility. Furthermore, there might have been an interaction between the words and the agent’s bizarre appearance, and this combined effect might have had an impact on memory.

An interesting aspect of the bizarreness effect concerns the kind of tests in which it occurs. Bizarre items are retrieved better in free recall than in cued recall tests [60]. This is in line with our results: the fact that participants learned better with the agent than with the human was evident in free recall tests [18]. Taking the above issues together, we therefore attribute our results to the bizarreness of Billie’s appearance and gestures. There might be an interaction between this factor and the unusual words in the artificial corpus used.

V. CONCLUSION

In our first experiment, participants had to acquire words in a foreign language [18]. The main concern was that they might not learn as well with the agent as with the human because of its avatar-like appearance and the robotic way in which it performs gestures. The results, however, demonstrated that the opposite was true. In a further step, a survey was conducted to investigate how participants perceived the quality of the gestures and the personality of both trainers. They did not provide us with consistent correlations to the first study, however. This might be because of the low number of participants (18) and the low statistical power deriving from small populations. On the other hand, the ratings in the survey can be interpreted as showing that the agent is simply perceived differently than the human. We have summarized the aspects of this difference and designated it as bizarreness, which might be the factor leading to the memory enhancement. The participants were not asked directly whether they perceived the agent as bizarre, but in future research we intend to develop a questionnaire capable of determining bizarreness. Furthermore, we believe that brain imaging designed to investigate how the human and the agent are perceived by the emotional and attentional networks of the brain might provide us with deeper insights into the learner-agent interaction.

REFERENCES

- [1] Heidig, S., and Clarebout, G. (2011), “Do pedagogical agents make a difference to student motivation and learning?”, *Educational Research Review*, 6(1), 27-54. doi: <http://dx.doi.org/10.1016/j.edurev.2010.07.004>
- [2] Clarebout, G., Elen, J., Johnson, W. L., and Shaw, E. (2002), “Animated Pedagogical Agents: An Opportunity to be Grasped?”, *Journal of Educational Multimedia and Hypermedia*, 11(3), 267-286.
- [3] Moreno, R., Mayer, R., and Lester, J. (2000), “Life-Like Pedagogical Agents in Constructivist Multimedia Environments: Cognitive Consequences of their Interaction,” paper presented at the World Conference on Educational Multimedia, Hypermedia and Telecommunications 2000. <http://www.editlib.org/p/16160>
- [4] Cassell, J., and Thorisson, K. R. (1999), “The power of a nod and a glance: envelope vs. emotional feedback in animated conversational agents,” *Applied Artificial Intelligence*, 13(4), 519-538.

- [5] Johnson, A. M., Ozogul, G., Moreno, R., and Reisslein, M. (2013), "Pedagogical Agent Signaling of Multiple Visual Engineering Representations: The Case of the Young Female Agent," *Journal of Engineering Education*, 102(2), 319-337. doi: 10.1002/jee.20009
- [6] Höck, K., Persson, P., and Sjölander, M. (2000), "Evaluating users' experience of a character-enhanced information space," *AI Communications*, 13(3), 195-212.
- [7] Lester, J. C., Converse, S. A., Kahler, S. E., Barlow, S. T., Stone, B. A., and Bhogal, R. S. (1997), "The persona effect: affective impact of animated pedagogical agents," paper presented at the Proceedings of the ACM SIGCHI Conference on Human factors in computing systems, Atlanta, Georgia, USA.
- [8] Cassell, J. (2000), *Embodied conversational agents*, Cambridge, Mass. ; London: MIT Press.
- [9] Baylor, A., and Ryu, J. W. (2003), "Does the presence of image and animation enhance pedagogical agent persona?", *Journal of Educational Computing Research*, 28(4), 373-395.
- [10] Beun, R. J., De Vos, E., and Witteman, C. (2003), *Embodied Conversational Agents: Effects on memory performance and anthropomorphisation*.
- [11] Mayer, R. E., and DaPra, C. S. (2012), "An embodiment effect in computer-based learning with animated pedagogical agents," *J Exp Psychol Appl*, 18(3), 239-252. doi: 10.1037/a0028616
- [12] Moundridou, M., and Virvou, M. (2002), "Evaluating the persona effect of an interface agent in a tutoring system," *Journal of Computer Assisted Learning*, 18(3), 253-261. doi: 10.1046/j.0266-4909.2001.00237.x
- [13] Mulken, S., André E., and Müller, J. (1998), "The Persona Effect: How Substantial Is It?", in H. Johnson, L. Nigay and C. Roast (Eds.), *People and Computers XIII* (pp. 53-66): Springer London.
- [14] Miksatko, J., Kipp, K., and Kipp, M. (2010), "The Persona Zero-Effect: Evaluating Virtual Character Benefits on a Learning Task with Repeated Interactions," in J. Allbeck, N. Badler, T. Bickmore, C. Pelachaud and A. Safonova (Eds.), *Intelligent Virtual Agents* (Vol. 6356, pp. 475-481): Springer Berlin Heidelberg.
- [15] Kim, Y., and Baylor, A. (2006), "Pedagogical Agents as Learning Companions: The Role of Agent Competency and Type of Interaction," *Educational Technology Research and Development*, 54(3), 223-243. doi: 10.1007/s11423-006-8805-z
- [16] van Daal, T., Donche, V., and De Maeyer, S. (2013), "The Impact of Personality, Goal Orientation and Self-Efficacy on Participation of High School Teachers in Learning Activities in the Workplace," *Vocations and Learning*, 1-20.
- [17] Webb, N. M. (2009), "The teacher's role in promoting collaborative dialogue in the classroom," *British Journal of Educational Psychology*, 79(1), 1-28.
- [18] Bergmann, K., and Macedonia, M. (2013), "A Virtual Agent as Vocabulary Trainer: Iconic Gestures Help to Improve Learners' Memory Performance," in R. Aylett, B. Krenn, C. Pelachaud and H. Shimodaira (Eds.), *Intelligent Virtual Agents* (Vol. 8108, pp. 139-148): Springer Berlin Heidelberg.
- [19] Macedonia, M., and Von Kriegstein, K. (2012), "Gestures Enhance Foreign Language Learning," *Biolinguistics*, 6(3-4), 393-416.
- [20] Larsen-Freeman, D., and Anderson, M. (2011), *Techniques and Principles in Language Teaching* (3rd ed. ed.), Oxford: Oxford University Press.
- [21] Zimmer, H. D. (2001), *Memory for Action: A Distinct Form of Episodic Memory?*, Oxford: Oxford University Press.
- [22] Engelkamp, J., and Krumnacker, H. (1980), "Imaginale und motorische Prozesse beim Behalten verbalen Materials," *Zeitschrift für experimentelle und angewandte Psychologie*(27), 511-533.
- [23] Engelkamp, J., and Zimmer, H. D. (1984), "Motor program information as a separable memory unit," *Psychological Research*(46), 283-299.
- [24] Macedonia, M., Müller, K., and Friederici, A. D. (2011), "The impact of iconic gestures on foreign language word learning and its neural substrate," *Human Brain Mapping*, 32(6), 982-998. doi: 10.1002/hbm.21084
- [25] Masumoto, K., Yamaguchi, M., Sutani, K., Tsuneto, S., Fujita, A., and Tonoike, M. (2006), "Reactivation of physical motor information in the memory of action events," *Brain Res*, 1101(1), 102-109. doi: S0006-8993(06)01399-0 [pii]10.1016/j.brainres.2006.05.033
- [26] Kelly, S. D., McDevitt, T., and Esch, M. (2009), "Brief training with co-speech gesture lends a hand to word learning in a foreign language," *Language and Cognitive Processes*, 24(2), 313 - 334. doi: 10.1080/01690960802365567
- [27] Saltz, E., and Donnenwerthnolan, S. (1981), "Does Motoric Imagery Facilitate Memory for Sentences - a Selective Interference Test," *Journal of Verbal Learning and Verbal Behavior*, 20(3), 322-332.
- [28] Craik, F. I. M. (2002), "Levels of processing: past, present, and future?", *Memory*, 10(5-6), 305-318. doi: 10.1080/09658210244000135
- [29] Backman, L., Nilsson, L. G., and Nouri, R. K. (1993), "Attentional Demands and Recall of Verbal and Color Information in Action Events," *Scandinavian Journal of Psychology*, 34(3), 246-254.
- [30] Knopf, M. (1992), "Gedächtnis für Handlungen. Funktionsweise und Entwicklung," unpublished post-doctoral thesis, University of Heidelberg.
- [31] Kormi-Nouri, R. (1995), "The nature of memory for action events: An episodic integration view," *European Journal of Cognitive Psychology*(7), 337-363.
- [32] Macedonia, M., and Knösche, T. R. (2011), "Body in Mind: How Gestures Empower Foreign Language Learning," *Mind, Brain, and Education*, 5(4), 196-211. doi: 10.1111/j.1751-228X.2011.01129.x
- [33] Macedonia, M. (2003), "Sensorimotor enhancing of verbal memory through 'Voice Movement Icons' during encoding of foreign language (German: Voice Movement Icons. Sensomotorische Encodierungsstrategie zur Steigerung der quantitativen und qualitativen Lerneffizienz bei Fremdsprachen)"(PhD Thesis), University of Salzburg, Salzburg.

- [34] Kopp, S. (2006), "Surface Realization of Multimodal Behavior from XML Representations in MURML," paper presented at the Workshop on Multimodal Behavior Generation, Reykjavik.
- [35] Salem, M., Kopp, S., Wachsmuth, I., and Joubin, F. (2009), "Towards Meaningful Robot Gesture," in H. Ritter, G. Sagerer, R. Dillmann and M. Buss (Eds.), *Human Centered Robot Systems* (vol. 6, pp. 173-182): Springer Berlin Heidelberg.
- [36] Bergmann, K., Kopp, S., and Eyssel, F. (2010), "Individualized Gesturing Outperforms Average Gesturing – Evaluating Gesture Production in Virtual Humans," in J. Allbeck, N. Badler, T. Bickmore, C. Pelachaud and A. Safonova (Eds.), *Intelligent Virtual Agents* (vol. 6356, pp. 104-117): Springer Berlin Heidelberg.
- [37] Fiske, S. T., Cuddy, A. J., and Glick, P. (2007), "Universal dimensions of social cognition: warmth and competence," *Trends Cogn Sci*, 11(2), 77-83. doi: S1364-6613(06)00329-9 [pii]10.1016/j.tics.2006.11.005
- [38] Hartmann, B., Mancini, M., and Pelachaud, C. (2006), "Implementing Expressive Gesture Synthesis for Embodied Conversational Agents," in S. Gibet, N. Courty and J.-F. Kamp (Eds.), *Gesture in Human-Computer Interaction and Simulation* (vol. 3881, pp. 188-199): Springer Berlin Heidelberg.
- [39] Gallese, V., Fadiga, L., Fogassi, L., and Rizzolatti, G. (1996), "Action recognition in the premotor cortex," *Brain*, 119(2), 593-609. doi: 10.1093/brain/119.2.593
- [40] Gazzola, V., Rizzolatti, G., Wicker, B., and Keysers, C. (2007), "The anthropomorphic brain: the mirror neuron system responds to human and robotic actions," *NeuroImage*, 35(4), 1674-1684. doi: S1053-8119(07)00096-1 [pii]10.1016/j.neuroimage.2007.02.003
- [41] Oberman, L. M., McCleery, J. P., Ramachandran, V. S., and Pineda, J. A. (2007), "EEG evidence for mirror neuron activity during the observation of human and robot actions: Toward an analysis of the human qualities of interactive robots," *Neurocomputing*, 70(13-15), 2194-2203. doi: http://dx.doi.org/10.1016/j.neucom.2006.02.024
- [42] Chaminade, T., Franklin, D. W., Oztup, E., and Cheng, G. (2005, 19-21 July 2005), "Motor interference between Humans and Humanoid Robots: Effect of Biological and Artificial Motion," paper presented at the Development and Learning, 2005, proceedings. The 4th International Conference on.
- [43] Chaminade, T., Hodgins, J., and Kawato, M. (2007), "Anthropomorphism influences perception of computer-animated characters' actions," *Social Cognitive and Affective Neuroscience*, 2(3), 206-216. doi: 10.1093/scan/nsm017
- [44] Press, C., Gillmeister, H., and Heyes, C. (2007), "Sensorimotor experience enhances automatic imitation of robotic action," *Proc Biol Sci*, 274(1625), 2509-2514. doi: F241417W014330VJ [pii]10.1098/rspb.2007.0774
- [45] Bisio, A., Stucchi, N., Jacono, M., Fadiga, L., and Pozzo, T. (2010), "Automatic versus voluntary motor imitation: effect of visual context and stimulus velocity," *PLoS ONE*, 5(10), e13506. doi: 10.1371/journal.pone.0013506
- [46] van Elk, M., van Schie, H. T., and Bekkering, H. (2011), "Imitation of hand and tool actions is effector-independent," *Exp Brain Res*, 214(4), 539-547. doi: 10.1007/s00221-011-2852-3
- [47] Walters, M. L., Syrdal, D. S., Dautenhahn, K., Ren, Boekhorst, T. (2008), "Avoiding the uncanny valley: robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion," *Auton. Robots*, 24(2), 159-178. doi: 10.1007/s10514-007-9058-3
- [48] Chaminade, T., Zecca, M., Blakemore, S. J., Takanishi, A., Frith, C. D., Micera, S. (2010), "Brain response to a humanoid robot in areas implicated in the perception of human emotional gestures," *PLoS ONE*, 5(7), e11577. doi: 10.1371/journal.pone.0011577
- [49] Dubal, S., Foucher, A., Jouvent, R., and Nadel, J. (2011), "Human brain spots emotion in non humanoid robots," *Social Cognitive and Affective Neuroscience*, 6(1), 90-97. doi: 10.1093/scan/nsq019
- [50] Krombholz, A., Schaefer, F., and Boucsein, W. (2007), "Modification of N170 by different emotional expression of schematic faces," *Biol Psychol*, 76(3), 156-162. doi: S0301-0511(07)00118-4 [pii]10.1016/j.biopsycho.2007.07.004
- [51] Bradley, M. M., Hamby, S., Low, A., and Lang, P. J. (2007), "Brain potentials in perception: picture complexity and emotional arousal," *Psychophysiology*, 44(3), 364-373. doi: PSYP520 [pii] 10.1111/j.1469-8986.2007.00520.x
- [52] Mori, M. (1970), "Bukiminotani [The Uncanny Valley]," *Energy*, 7, 33-35.
- [53] Steckenfinger, S. A., and Ghazanfar, A. A. (2009), "Monkey visual behavior falls into the uncanny valley," *Proceedings of the National Academy of Sciences*, 106(43), 18362-18366. doi: 10.1073/pnas.0910063106
- [54] Riefer, D. M., and Rouder, J. N. (1992), "A Multinomial Modeling Analysis Of The Mnemonic Benefits Of Bizarre Imagery," *Memory and Cognition*, 20(6), 601-611. doi: 10.3758/bf03202710
- [55] Craik, F. I. M., and Lockhart, R. S. (1972), "Levels of Processing - Framework for Memory Research," *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671-684.
- [56] Nicolas, S., and Gounden, Y. (2011), "Bizarre imagery and memory," *Psychologie Francaise*, 56(4), 203-208. doi: 10.1016/j.psfr.2011.10.002
- [57] Gounden, Y., and Nicolas, S. (2012), "The impact of processing time on the bizarreness and orthographic distinctiveness effects," *Scandinavian Journal of Psychology*, 53(4), 287-294. doi: 10.1111/j.1467-9450.2012.00945.x
- [58] Engelkamp, J., Zimmer, H. D., and Biegelmann, U. E. (1993), "Bizarreness effects in verbal tasks and subject-performed tasks," *European Journal of Cognitive Psychology*, 5(4), 393 - 415.
- [59] Geraci, L., McDaniel, M., Miller, T., and Hughes, M. (2013), "The bizarreness effect: evidence for the critical influence of retrieval processes," *Memory and Cognition*, 1-10. doi: 10.3758/s13421-013-0335-4
- [60] Einstein, G., and McDaniel, M. (1987), "Distinctiveness and the Mnemonic Benefits of Bizarre Imagery," In M. McDaniel and M. Pressley (Eds.), *Imagery and Related Mnemonic Processes* (pp. 78-102): Springer New York.

ONLINE SURVEY

Questions asked about the agent's gestures

1. Does the gesture fit well the word's meaning?
2. Does the gesture look natural?
3. Does the gesture look fluent?
4. Is the speed of execution of the gesture appropriate?

Questions on the trainers' personality and gestures

5. Is the virtual agent pleasant?
6. Is the virtual agent friendly?
7. Is the virtual agent reliable?
8. Is the virtual agent intelligent?
9. Is the virtual agent competent?
10. Is the virtual agent human-like?
11. Is the virtual agent sympathetic?

Questions about the trustworthiness of the trainer

12. Imagine the agent as (your) teacher. Would you trust him in this function?