An Empirical Study on the Trustworthiness of Life-Like Interface Agents

Susanne van Mulken, Elisabeth André, and Jochen Müller
DFKI GmbH, Stuhlsatzenhausweg 3, D-66123, Saarbrücken, Germany

1 Introduction

In this paper, we describe an experiment investigating the impact of the personification of interface agents as life-like characters on the agents' trustworthiness. Current research in intelligent interfaces moves strongly toward interfaces in which the communication between human and computer is mediated by intelligent interface agents---programs that may be authorized to autonomously act upon the user's behalf. A prerequisite for the acceptance of this new form of HCI is that the user trust the agent and indeed delegate tasks to it. Besides on factors such as a cost-benefit tradeoff and the possibility for the user to effectively communicate the task to the agent, delegation depends on the trustworthiness of the agent. Trustworthiness in turn may depend on the agent's underlying behaviour (e.g., the agent's competence and its perceived predictability) and on the way the agent presents itself to the user (e.g., Milewsky & Lewis 1997).

Related research suggests that it may be advantageous to make interface agents appear more human-like. For instance, there is evidence that when users interact with computer-based systems, they are biased at first toward distrust, whereas in interaction with other humans, there is a tendency toward trust (see Muir 1987). In addition, some empirical studies (e.g., Sproull et al. 1994) provide evidence that visually personifying the interface (e.g., through a computer-animated face) can lead to general social facilitation. Finally, facial expressions, body gestures, and vocal intonations are interpreted in terms of trustworthiness and can also be used to express trustworthiness (see DePaulo 1992).

A more specific line of research in intelligent interfaces concentrates precisely on personifying the look of interface agents: Agents are personified by means of life-like characters which inhabit the user interface.

The questions investigated in the current study are: How big is the impact of a life-like character on the trustworthiness of the system compared to that of other

ways of representation, such as text or audio information? Furthermore, To what extent does the quality of the underlying system behaviour influence any possible impact?
Our hypothesis was that agents represented by life-like characters, being more human-like, should be found more trustworthy than those represented by text or audio information. Furthermore, we expected this effect to show up only in case of low-quality system behaviour; in case of high-quality system behaviour, we expected to find ceiling effects.

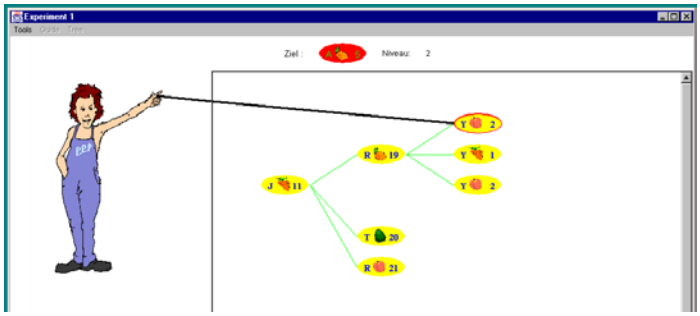

Figure 1: Navigation environment with Persona assistance.

2 Approach

In our study we used two life-like interface agents developed at DFKI (the PPP Persona and a video agent). These agents are developed to present multimedia material to the user following the directives of a script (see André, Rist, & Müller 1998). This material is either automatically generated or retrieved from the World-Wide Web. While the user views the presentation, the agents can among other things comment on particular parts and highlight them through pointing gestures.
In order to measure the impact on trustworthiness, we designed a (Web-like) navigation task similar to that used by St. Amant & Dulberg (1998). The subjects' task was to navigate through an expandable tree structure in search for a goal node, while minimizing the number of expansions (see Figure 1). The nodes in the tree were expandable by clicking on them. The subjects were assisted in their search either by the Persona (a cartoon-style agent), by the video agent, by audio information, or by textual information. The assistance consisted of the system's giving recommendations as to where to proceed with the search (i.e., what node to click on next). Prior to the experiment, the subjects were given

- the information that the system's assistance would not always be correct (i.e., it might guide the user to a wrong path through the tree); and

- information about the structure of the tree (e.g., characteristics of the nodes' labels, the relationships between parent and child nodes, and the level at which the goal node was located).

Thus, at each navigation point, the subject had to decide whether to rely on the system's assistance or rather choose his or her own path through the tree.

3 Design

As independent variables we defined the Representation of the Agent (RoA) (with four levels) and Quality of the Agent (QoA) (with two levels). As dependent variables we chose the ratio of the number of clicks on recommended nodes and number of expansions, and subjective ratings of trustworthiness on a questionnaire. RoA was manipulated within subjects, whereas QoA was manipulated between subjects. Thus, the experiment had a 4 (RoA) x 2 (QoA) mixed design.

Representation of Agents. The representation of the agents could take on four values: Text, Audio, Video-Agent or Persona. In case of Text, recommendations were given in text, accompanied by a blinking of the intended nodes. In the Audio condition, the recommendations were given auditorily, again accompanied by blinking. In the case of Persona, the recommendations were given by the Persona cartoon who, besides speaking the recommendation, pointed with a pointer stick at the intended nodes. The Video-Agent condition differed from Persona only in that the agent was represented by a real-looking person.

Quality of Agents. The quality of the recommendations could take on two values: high or low. In the case of high-quality assistance, the probability of system's recommendations being correct was .75, whereas with low-quality assistance, this probability was .25.

Subjects were allocated to the QoA condition randomly. The order of RoA conditions was counterbalanced. Furthermore, so as to control for the particular voice, we used both a normal pitch voice and a low-pitch voice.

4 Procedure

The experiment started with a test session, consisting of a sequence of four test trees. The aim of this first test session was to have the subjects grow acquainted with the environment and get a feeling of how the nodes in the trees were organized. Here, the system would not give any recommendations at all. After this first test session, the subjects would proceed with a second test session. Here, the system would give recommendations by simply encircling the nodes it recommended. The aim of this session was to have the subjects experience the quality of the recommendations. After the test sessions, the sequence of experimental sessions with either Text, Audio, Persona, or Video-Agent

recommendations would begin. At the end of the experiment, the subjects were given a questionnaire asking them to rate the trustworthiness of the different agent versions.

5 Results and Discussion

Proportion of recommended expansions made. For the objective data, we found a main effect of QoA ($F(1,30)=34,64$; $p<.01$) As expected, in case of low-quality assistance, subjects followed the recommendations less often than in case of high-quality assistance. In addition, in case of low QoA, the proportion of recommended expansions actually made dropped off going from the Video- and Cartoon-agents to Audio and to Text. However, this effect was not statistically significant. Finally, we found an interaction between QoA and RoA ($F(3,90)=3,02$; $p<.05$). This interaction was due to a decrease in the number of recommended expansions for Audio and Text in the case of low QoA, and an increase for these forms of representation in case of high QoA (see Figure 2).

Subjective ratings of trustworthiness. For the subjective data, we found a main effect of RoA ($F(3,90)=2.71$; $p<.05$). In case of high-quality assistance, the trustworthiness of the agent represented by Text was rated higher than that of the agents. For the low-quality assistance condition, there were no significant differences between the ratings. In addition, there was no significant effect of QoA (see Figure 3).

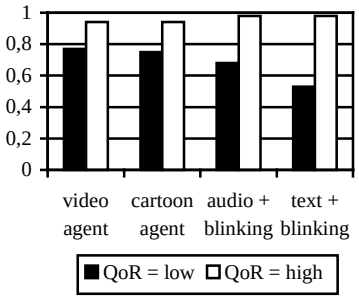

Figure 2: Proportions of recommended expansions

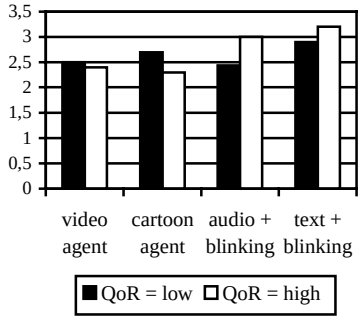

Figure 3: Trustworthiness ratings

Contrary to our expectations, we did not find any significant differences between the ways in which the agents were represented with regard to the proportion of expansions of recommended nodes. One of the reasons that may account for this is that our cartoon and video agents might have not been life-

like or realistic enough. For instance, although the lip movements of the agents were synchronised, they were restricted in variability. Furthermore, the agents' body movements were not always optimally smooth. Improving this might raise the agents' trustworthiness and enlarge the numerical differences we found.

Considering the subjective ratings, there appears to be a slight dissociation between the subjects' opinions concerning the trustworthiness of the respective assistance versions and their behaviour: Although subjects in the high-quality assistance condition indicate to trust the textual assistance more than the personified agents, there was no reflection of this in their behaviour.

6 Conclusions

The experiment presented in this paper provides support for the hypothesis that in the trustworthiness of interface agents, the competence of the system plays a major role. Furthermore, we found numerical differences in how the agents were represented, with the trustworthiness of personified agents being higher than those of the unpersonified ones (on the behavioural measure). However, as these differences did not reach statistical significance, we conclude that personification in itself does not appear to be a sufficient condition for raising trustworthiness. Further studies are necessary to investigate what attributes of personified interface agents contribute most significantly to their trustworthiness.

7 References

André, E., Rist, T., and Müller, J. (1998). Integrating Reactive and Scripted Behaviors in a Life-Like Presentation Agent. In *Proc. 2nd Int. Conference. on Autonomous Agents*, ACM Press, pp. 261-268.

Milewsky, A.E. and Lewis, S.H. (1997). Delegating to software agents. *Int. J. Human-Computer Studies*, 46, 485-500.

Muir, B.M. (1987). Trust between humans and machines, and the design of decision aids. *Int. J. Human-Computer Studies*, 27, 527-539.

St. Amant, R. and Dulberg, M.S. (1998). An experiment with navigation and intelligent assistance. In *Proc. Intelligent User Interfaces'98*, ACM Press, pp. 171-178.

DePaulo, B.M. (1992). Nonverbal Behavior and Self-Presentation. *Psychological Bulletin*, 111 (2), 203-243.

Sproull, L., Subramani, M., Kiesler, S., Walker, J.H., and Waters, K. (1996). When the Interface Is a Face. *Human-Computer Interaction*, 11, 97-124.