Be Quiet? Evaluating Proactive and Reactive User Interface Assistants

Jun Xiao, Richard Catrambone & John Stasko

College of Computing / School of Psychology & GVU Center
Georgia Institute of Technology
Atlanta, GA 30332

junxiao@cc.gatech.edu, rc7@prism.gatech.edu, stasko@cc.gatech.edu

Abstract: This research examined the ability of an anthropomorphic interface assistant to help people learn and use an unfamiliar text-editing tool, with a specific focus on assessing proactive assistant behavior. Participants in the study were introduced to a text editing system that used keypress combinations for invoking the different editing operations. Participants then were directed to make a set of prescribed changes to a document with the aid either of a paper manual, an interface assistant that would hear and respond to questions orally, or an assistant that responded to questions and additionally made proactive suggestions. Anecdotal evidence suggested that proactive assistant behavior would not enhance performance and would be viewed as intrusive. Our results showed that all three conditions performed similarly on objective editing performance (completion time, commands issued, and command recall), while the participants in the latter two conditions strongly felt that the assistant's help was valuable.

Keywords: agent, intelligent assistant, anthropomorphism, empirical study, text editing, help, evaluation, interface agent

1 Introduction

The use of agent-based characters or anthropomorphic/personified assistants in user interfaces is ripe with controversy (Laurel, 1990). Strong arguments both questioning (Lanier, 1995; Shneiderman, 1997) and supporting (Cassell, 2000; Laurel, 1990) this approach in user interface design have been voiced, but relatively little careful empirical study has been done (Dehn & van Mulken, 1999; Sanders & Scholtz, 2000). As Isbister and Doyle (2002) state, "Rigorous evaluations of benefits to the user are rare, and even when performed are subject to considerable criticism owing to the difficulty of finding objective measures of success."

Perhaps one reason for the lack of empirical study is the breadth of this area. Research on agents in user interfaces typically focuses on one of three different goals: Improving our understanding of human-to-human interactions to inform agent design, creating agent architectures and systems, or evaluating agent-based interfaces and systems (Pelachaud & Poggi, 2001). We feel that the evaluation of agent systems, the third component, has received less focus than the other two components. This deficiency has been noted by others, and recently, frameworks for further evaluation work have been proposed (Isbister & Doyle, 2002; Ruttkay, Dormann, & Noot, 2002).

One challenging aspect to empirical study is that agents can serve many different functions and take

on many different forms. For example, agents can serve as avatars, guides, intelligent assistants, or entertainers. Agents also can be disembodied, embodied, personified, anthropomorphic, 2D cartoons, 3D video depictions, or some combination.

Our overall research program involves a systematic study of the factors influencing agent success. These factors are based on prior research and considerations about human performance. In our research, we study the utility of an agent acting as an assistant, helping users perform tasks. The agent is presented through a 3D "talking head" face on the display. In the present study, we specifically examined two questions: How do people perform with the assistant's help as compared to other, more traditional help techniques, and how will proactive assistant behavior affect performance and be perceived by users. By "proactive," we mean that the assistant, in addition to answering queries, will make unsolicited suggestions.

The specific task in focus is that of learning and using the operations of a new software system. In a manner similar to an approach taken by Mackay (2001), we introduced users to an unfamiliar text editing tool and required them to make changes to a marked-up document.

The traditional form of assistance is some form of documentation or help manual. We utilized a printed set of help pages documenting the key combinations for different commands as the control condition in our study. The two other conditions involved an interface agent, presented as a 3D human face, that a user interacts with and queries.

We measured objective user performance via the time to complete editing, the number of editing commands issued, and a post-editing session quiz on the key sequences for all the editing commands. We also elicited subjective user views on the difficulty of the task and the utility of the agent in the two conditions. Additionally, we queried participants about other agent attributes (e.g., intelligence) and we attempted to assess participant personality characteristics and examine whether they correlated with any perceptions of the agent.

Presumably, written help in the form of documentation is the "state-of-the-art" for assisting the task we examined. Accordingly, we hypothesized that condition would result in top objective participant performance. One goal of our study was to examine how well the two agent conditions would compare to this standard.

Another objective of the study was to analyze proactive agent behavior. The best-known example of similar agent behavior is likely that of the Microsoft Office Paper-Clip assistant "Clippy" that many computer users have encountered (Schaumburg, 2001). Strong anecdotal evidence coupled with interview results from a prior experiment we performed (Catrambone, Stasko, & Xiao 2002) suggest that a majority of people in the user demographic group we studied have a negative view of the Office assistant character. Frequently, people cite the proactive suggestions as a particularly problematic attribute of the assistant.

One of our goals was to examine whether the simple act of the assistant making suggestions (interrupting, in some sense) caused this reaction or whether it was the quality of the suggestions that mattered most.

In the next section, we review prior related work that informs research in this area.

2 Related Work

The creation of interface assistants is a complex endeavor that requires research in a variety of areas including artificial intelligence, computer animation, interface design, sociology, and psychology. Also, a wide variety of factors might influence how the assistants are viewed (Catrambone et al., 2002). In this section, we will focus on studies of proactive interface assistants. For reviews of general evaluation studies of interface agents, see (Catrambone et al., 2002; Dehn et al., 1999; Ruttkay et al., 2002).

The effects of proactive interface assistants have been studied in a variety of user task settings such as interactive learning, web browsing, document composition and real estate search (Cassell et al., 1999; Fischer, Lemke, & Schwab 1985; Lester et al., 1997; Lieberman, 1997; Rhodes, 2000). Lester et al. (1997), for example, designed an animated pedagogical agent, Herman the Bug, who inhabited a learning environment and proactively provided advice while the student designed plants. An empirical study showed that such an assistant could increase both the learning performance and the student's motivation. The active communicative behaviors had a strong positive effect on learners' perception of their learning experience, resulting in a more attentive, more engaged audience.

Proactive help also has been investigated in software education situations. The COACH system assisted learners of the Lisp programming language by giving a user feedback about the user's knowledge and predicting the user's future goals and

actions (Selker, 1994). The system monitored user actions to create a user model, which was used to decide when and what kind of unsolicited advice would be offered. The usability study conducted to examine the COACH system with and without the proactive help showed that users in the proactive help group on average utilized all available materials, felt more comfortable with Lisp, had higher morale, and wrote five times as many functions than users without the proactive help.

The fact that a computer assistant measures people's performance and passes judgment, however, may cause some users to perceive this sort of proactive help as intrusive and offensive rather than helpful. Rickenberg and Reeves (2000) found a higher level of anxiety and more mistakes in the solution of a task among users with a monitoring computer character, compared to those with a character that ignored them. Their results also indicated that these effects depend on the personality traits of the users.

As software systems become more complex and functionally rich, end-users' needs for assistance seem likely to amplify. User interface assistants, in one form or another, are a potential technique for addressing this problem. The studies cited above and others, however, suggest that progress in this area has been hampered by a lack of empirical studies to guide the design and construction of such assistants.

3 Experiment

3.1 Overview

This experiment was designed to examine the effect of an interface assistant's initiative on people's performance relative to a more traditional manual, and to examine people's perceptions of the assistant. A secondary goal was to examine whether a user's personality plays a role in his or her reactions to reactive and proactive agents. An editing task was designed to represent an opportunity to use an interface assistant primarily as a replacement for a manual. Each participant was instructed to complete the editing task, working with a proactive assistant, a reactive assistant, or a manual that listed all the commands.

We employed a Wizard of Oz methodology to remove agent ability as a variable, although participants believed that they were interacting with a fully functional help system. Such an approach is commonly used when the technology for building the interface does not yet exist or is not yet sophisticated enough to be applied to real applications (Dahlback, Jonsson, & Ahrenberg, 1993). This technique allowed us to focus on evaluating the proactivity of an interface assistant constructed to be consistent with certain constraints, instead of the quality of the supporting help system. In some sense it yielded an upper-bound estimate of the performance that was attainable by an interface assistant that understood users and offered help.

3.2 Goals

So far, little research has been conducted to investigate the effect of proactive interface assistants. This is partly due to the fact that users' perceptions of an interface assistant and their task performance are strongly affected by the competence of the supporting software system and the quality of the replies and suggestions made by the interface assistant. While a set of experiments could examine how differing levels of competence affect user performance and impression, there is a fundamental question that needs to be addressed first: Will people's performance be enhanced and will they like to work with a proactive interface assistant if it is actually competent?

If a person could seek aid from a smart, spoken natural language help system, would that be better than an on-line reference manual? Is unsolicited advice from an assistant intrusive and frustrating to users, or welcomed by users? We believe that these questions should be answered on the basis of scientific evidence rather than personal anecdote and experience. In this study, we were particularly interested in evaluating the usefulness of such proactive interface assistants via both the performance and satisfaction dimensions. We would like to understand why people disable, far more often than use, interface assistants such as the Microsoft Paper Clip. Is it because personified proactive interface assistants are fundamentally a bad idea, this particular assistant was badly designed, the assistant was chosen for the wrong domain, or an interaction of those factors?

We hypothesized that user reactions to an interface assistant would vary as a function of the initiative of the assistant. A synthetic character that makes unsolicited suggestions might lead the user to feel that it was annoying, intrusive, distracting, and less helpful compared to an interface assistant that responded only to users' questions.

However, it is possible that the objective performance of users might be affected in a different way due the initiative of an interface assistant. A proactive assistant might help users to perform their

tasks more effectively and learn better about the software system. This might occur because of the timing of the proactive help. Timely help could help a learner to remember the information. This would occur because the information is more likely to be consolidated with the current goal because the learner is not distracted by having to form a request for information This reduction in cognitive load has proven beneficial in a variety of learning situations (Mayer & Anderson, 1992; Mousavi, Low, & Sweller, 1995).

Finally, we expected that personality differences might lead to different reactions to the proactive behavior of an interface assistant by users, as suggested by Nass, Steuer, & Tauber (1994), but objective performance of users would likely not be affected by personality attributes. For instance, someone who is extroverted may find a proactive agent more friendly than one that responds only when asked a question.

3.3 Method

Participants

Forty-nine Georgia Tech undergraduates participated for course credit and were randomly assigned to conditions. Participants had a variety of majors and computer-experience backgrounds. Analysis of the demographic information collected during the experiment showed that computer experience backgrounds and gender were approximately balanced across conditions.

Software and Equipment

In order to make the text editor unfamiliar to all participants, we built our own text editor (see Figure 1). There were no menus or other interface objects besides the text window itself, and the mouse was disabled in the editor environment. To issue an editing command, the participant had to press a specific key combination. The editor offered a rich set of commands for higher-level operations on words, lines, and sentences besides basic character-oriented commands.

The interface assistant used in the experiment had an animated 3D female appearance (though somewhat androgynous) that blinked and moved its head occasionally in addition to moving its mouth in synchronization with the synthesized voice. However, no intonation, speech style, gaze patterns, and facial expressions were used to convey additional information. A prior study (Catrambone et al., 2002) suggested that people felt comfortable talking to a computer assistant with this appearance. The assistant stayed in a small window at the upper-right corner of a participant's computer screen.

Participants were run individually using a computer equipped with a microphone and speaker. The experimenter in an adjacent room monitored the editing activities of a participant via composite video images from two video cameras located in the participant's room and the participant's computer screenshot. The entire session was recorded and keystrokes issued by each participant were logged.

Figure 1: Interface that participants encountered.

Procedure

Demographic Information: Each participant filled out a demographic information sheet and a 50-item version of the "Big 5" personality inventory (McCrae & Costa, 1987).

Training: Participants first viewed a short video that described the various functions (e.g., copy, paste) and the specific key combinations needed to issue the commands. The editor interface and the usage of each command were shown and explained as a sample document was being edited in this tutorial video. Participants were asked to watch the video closely but they were informed that they would be able to get help from a computer assistant (or could look up commands in the *manual* condition) when carrying out editing tasks.

Editing Task: The editing task required participants to use the text editor to modify an existing document by making a set of prescribed changes such as deleting, inserting, and moving text. Pilot testing was conducted to ensure that the tasks were of appropriate difficulty and that the number of commands (18) was sufficiently large so that participants would be likely to need to consult the assistant or manual for help.

The body of text to be modified was shown on 25 pages of paper. On each page there was only one marked modification that the participant needed to perform. The next page showed the resulting text after the previous modification had been made and the next required modification. Participants were instructed to perform the tasks in the order of the pages and their ultimate goal was to revise the document as accurately and as quickly as they could.

The editing commands fell into two categories: those that involved moving around (e.g., move forward/backward by a word, move to beginning or end of sentence, mark beginning and end of some text) and those that altered text (e.g., delete a sentence, delete marked text, copy marked text, paste marked text). Tasks that required some of these commands were constructed so that early in the document a particular command (e.g., delete a sentence) would likely be used twice in succession with no intervening commands, while a third opportunity to use the command occurred towards the end of the document. In cases in which a participant either asked for help from the agent for a particular command or received proactive help, the second opportunity allowed the participant to immediately use the information. Additionally, in the proactive condition, the advice might also serve as a reminder that a particular command existed, thereby increasing the chance that the participant would use the command (perhaps rather than some less efficient command, if one existed, such as deleting a sentence character by character). The third opportunity simply functioned as a subsequent check of the participants' memory of the commands after some time had elapsed.

Conditions: The experiment was a three-group, between-subjects design. Each participant worked with either a proactive assistant, a reactive assistant, or was given a manual that listed all the commands.

1. Reactive Condition: Participants were informed that as they worked on the document, if they did not remember the key sequence for a command, they could ask the computer assistant. Before the session started, the experimenter asked the assistant a sample question, "How do I insert text into a document" to make sure that participants understood how the system worked.

As mentioned earlier, the interface assistant's responses were controlled through a Wizard of Oz technique. One experimenter was in the room with the participant to introduce the experimental materials, and a second experimenter (the wizard) was in an adjacent room, monitoring the questions asked by the participant. The wizard had full knowledge of all the editing commands. The assistant always answered questions in the following predefined manner: "To X press Y" (e.g., "To delete a character press control-d"). The wizard determined which answer best fit the participant's question A variety of responses covering other situations was also prepared. This included responses such as asking the participant to repeat the question or to state that

the assistant was not able to provide an answer (in cases in which a participant asked for a function that the editor did not possess). Participants were advised that the assistant's knowledge was restricted to the functionality of the text-editing tool.

2. Proactive Condition: In addition to what participants were told in the reactive condition, participants in the proactive condition were reminded that the assistant might sometimes offer them unsolicited help.

In order to make the proactive advice from the wizard resemble that of a competent computer assistant with respect to performance and consistency, strict rules were imposed on the wizard who observed participants' editing activities, reducing the potential for variability and bias.

First, if the wizard found that a participant performed a subtask inefficiently, for example, moving the cursor character-by-character to the beginning of the sentence instead of applying the move-to-the-beginning-of-a-sentence command, then the assistant would suggest a more efficient way to perform the operation. Second, if the assistant could predict what the next task would be, for example, a copy would have to be followed by paste operation, it would tell the participant the key combination needed for the very next task (e.g., it would tell the participant the paste command right after the participant had copied some text). Third, the assistant would make a suggestion about a command only if the participant had not used that command or previously asked about it. Fourth, each command could be proactively suggested at most once. Finally, the assistant did not explain the reasoning for its suggestions to the participant.

Pilot testing and a task analysis determined that five commands could be suggested by the proactive assistant: move cursor forward/backward by words, move cursor to the beginning/end of a sentence, delete the current sentence, mark the end of region to be cut or copied, and paste. While the wizard of course knew what tasks the participant was doing, the assistant was not supposed to be omniscient. Rather, the proactive advice came only in situations in which it was logically plausible to infer the next action given the current one (e.g., pasting some text after copying some text).

3. Manual Condition: Instead of working on the task with the help of an assistant, a set of pages that listed all the commands and the corresponding key sequences were available to participants in this condition during the test. The commands were

ordered in the same way as they were introduced in the tutorial video.

Post-test and Free Response Interview: After completing the editing tasks, participants were tested on their memory of the various editor commands with a form that listed all the commands. Participants wrote down the corresponding key sequences that they recalled and were allowed to look at the keyboard to help with recall. After that, participants filled out the personality inventory once again, but this time did it for the agent, that is, they rated the agent's "personality." Finally, participants filled out a questionnaire concerning issues such as the usefulness and friendliness of the agent (the items are listed in the Results section) and were asked a few questions about the assistant and related issues by the experimenter.

Measures

The dependent variables in the experiment were time to do the editing tasks, the responses to the items in the questionnaires, the replies to the questions posed by the experimenter, answers to the test on editing commands, and the personality inventory responses.. The questions posed by the experimenter were open-ended and provided participants an opportunity to give their impressions about the assistant's personality, helpfulness, and intelligence. With respect to editing task performance, besides time on task we counted how many keystrokes participants pressed in total to carry out all the tasks, and how many times participants queried or received help.

4 Results

4.1 Performance Measures

The time (in seconds) to do the editing task did not differ as a function of condition (proactive agent: 874.2, reactive agent: 869.2, manual: 805.0); $F(2, 46) = 0.53$, $MSE = 45289.85$, $p = .59$. The number of commands issued by participants while doing the editing task also did not differ as a function of condition (proactive: 730.3, reactive: 970.3, manual: 939.0); $F(2, 46) = 1.05$, $MSE = 255673.89$, $p = .36$. These numbers include all keystrokes, including those for cursor movement, except those involving text entry. Although the differences in means would seem to favor the proactive condition, the difference is not significant partly due to high within-group variability.

There was a significant difference among conditions in the number of times they received help

(proactive: 13.6, reactive: 8.6, manual: 8.4); $F(2, 46) = 5.88$, $MSE = 23.30$, $p < .006$). Pairwise comparisons showed that the proactive condition received more help than both the reactive and manual conditions (both p's $< .005$) while there was no difference between the reactive and manual conditions ($p = .91$). For participants in the reactive agent condition this number reflected the number of times they asked the agent a question. For participants in the proactive agent condition it reflected the number of times participants asked for help plus the number of times the agent proactively offered help. For the manual condition it reflected the number of times participants consulted the manual. Given that there was no difference among groups on task time, there did not appear to be a time cost associated with frequency of help.

Finally, there was no significant difference among the groups on the post-test on command recall (proactive: 12.1, reactive: 11.6, manual: 10.8); $F(2, 46) = 0.83$, $MSE = 8.23$, $p = .44$.

	Pro-active (*n*=15)	Re-active (*n*=18)	AVG	p-val.
Agent was:				
Worthwhile	1.8	3.2	2.6	*p < .05*
Intrusive	6.7	7.1	6.9	*p = .66*
Distracting	7.4	7.3	7.3	*p = .86*
Friendly	4.9	3.6	4.2	*p = .14*
Annoying	6.5	6.5	6.5	*p = .97*
Intelligent	3.8	3.2	3.5	*p = .48*
Competent	3.3	2.7	3.0	*p = .46*
Agent helped with task	1.6	2.8	2.3	*p = .09*
Like to have agent	3.3	3.7	3.5	*p = .68*
Editor diff. to use	6.1	5.7	5.9	*p = .71*
Made suggestions at right time	1.7	n/a	n/a	n/a

Table 1: Responses to Questionnaire Items as a Function of Type of Agent. Note**:** Responses were on a Likert scale from 1 (strongly agree) to 9 (strongly disagree).

4.2 Subjective Assessment

Table 1 shows the mean responses to the questionnaire items for the two agent conditions. There was an effect of agent type for only the question about whether the agent was worthwhile that favored the proactive condition. However, given the large number of statistical tests conducted, the safest conclusion is that there appears to be little

or no difference between the two agent conditions. It is interesting that for most items the average response tended to be on the positive side. That is, the agents were worthwhile and helpful, but not distracting, intrusive, or annoying. Participants in the proactive condition also strongly agreed that the agent made suggestions at the right time.

4.3 Personality Ratings

Participants' ratings of themselves and the agent were examined. The correlations between the self and agent scores on the dimensions were non-significant. There was also no relationship between agent condition and the personality ratings given to the agent (e.g., the proactive agent was not viewed as more extroverted than the reactive agent). Finally, there was no relationship between self-ratings and responses on the questionnaire about the agent (i.e., items in Table 1).

4.4 Interview Feedback

In the interview sessions with the participants in the agent conditions, we asked whether the agent was helpful, intelligent, and had a personality. We additionally asked people in the proactive agent group whether the unsolicited suggestions were helpful, and we asked people in the reactive agent group how they would have felt if the agent had made suggestions We asked for suggestions on how to improve the agent, and finally, we asked each if they were familiar with the Microsoft Office assistant Clippy, and if so, their opinion of it.

All but two participants in the agent conditions, both in the reactive condition, felt that the interface assistant was helpful. It is interesting to note that these two participants also had the two slowest completion times in the entire experiment. Subsequent examination revealed nothing unique with respect to demographics or personality of the two individuals.

The majority of participants felt that the agent did not have a personality, and impressions were mixed about the agent being intelligent. Here, one (reactive agent) person's response was quite interesting, "Yeah, she knew what she was talking about. She was more intellectual than me, a level above me."

In the reactive group, even though all but two participants felt the assistant was helpful, only four reacted positively when asked if they would like proactive behavior added. Here, their prior experiences (see below) may be influencing their reactions. Curiously though, these negative speculations contrasted strongly with the views of people encountering the proactive agent. With

respect to proactive behavior, 12 of the 15 participants had favorable reactions. Typical comments included, "It was actually pretty helpful," "(They were) timed right—great," "It answered questions before I had to ask," and "(It said what) was the next thing on my mind." One of the three negative views actually could be interpreted as a positive, "It didn't do enough." The other two included, "It might get annoying," and "(It was) a little obnoxious. When I really needed help, [the agent] didn't say anything."

As for improvements, participants suggested a variety of ideas, some contradictory. Six of the respondents in the proactive agent group suggested improving the quality of the agent's voice. One person wanted to see the whole person rather than just the face, and another stated, "It could've been more pleasant. It wasn't engaging." Four people in the reactive group suggested improving the voice. Three suggested removing the face, one suggested that it not be there all the time (only when answering), and one wanted a more animated and expressive appearance. One respondent wanted more functionality and another wanted better advice.

All participants indicated they were familiar with the Microsoft Office assistant Clippy, and their impressions were decidedly negative. The most frequent adjective used to describe Clippy was "annoying." Other comments included:

"It's not very helpful. You've got to stop what you're doing and click on it."

" I make it go away. It dances and is distracting."

"It doesn't pop up at the right times."

"It would be better if you asked it rather than typed."

"I don't like it at all. It's like a stalker."

Most notable to us is the striking dissimilarity in reactions to the proactive agent in the experiment as compared to the Office assistant.

5 Discussion

As noted in the introduction, evaluations of interface agents with respect to objective performance measures are relatively rare. In this study, people using the reactive and proactive assistants did not perform significantly differently than those using a very familiar form of help, the printed page. For one important measure, number of commands issued, the trend favored the people with the proactive assistant, with variability in performance across individuals in the groups eliminating a significant effect.

With respect to subjective assessment, participants using the assistants strongly indicated that they felt the interactions were helpful. The participants, on the whole, seemed to be comfortable with and took advantage of the spoken interactions with the assistants.

Critics of agent interactions may argue that the participants using the assistants did only as well as those using traditional printed documentation, even with an idealized Wizard-driven agent, one that is likely beyond current technological implementation possibilities. We would argue that to expect participants using the agent to outperform those using the manual, particularly in such a limited, straightforward task, would be unrealistic. In some sense, the fact that participants using the agents did just as well as those using a printed manual is significant. One might speculate that when a task becomes more complex, the benefits of spoken interaction with an agent as done here may increase. The key finding of this paper is that prior experiences and stereotypes have not soured people's views of proactive agents. When an agent provides useful, relevant assistance, the aid is appreciated.

In all likelihood, whether a person prefers printed, on-line, or assistant-based help will vary by individual. Particular people's background, experience, and personality will make them more or less comfortable with the different styles of interfaces and help.

One striking finding was the contrast between the views of proactive behavior between participants in the reactive and proactive conditions. The majority of participants using only the reactive agent felt that suggestions would not be welcome. In contrast, people receiving proactive advice almost uniformly felt it was helpful. We speculate that prior experiences with interface agents may be biasing the views of the participants in the reactive condition. The strong, negative reactions toward the Microsoft Office assistant, as noted above, seem to support this view.

One conclusion we draw from the study is that proactive suggestions in and of themselves are not necessarily problematic. Suggestions that are understandable and appropriate can be well received. Researchers must work to improve techniques for automated analysis of user goals and intentions, and then design interface agents that suggest actions only when confidence is extremely high that the advice will be worthwhile.

In prior work, Horvitz (1999) presented a set of principles for determining the right moment to interact with the user. He adopted a decision theoretic approach to regulate the behavior of interactive assistants designed to provide unsolicited help and balance the utility of the information with the cost of disruption it would cause the users.

Mackay (2001), however, pointed out a different solution to the same problem. She found that subjects learned commands for a text editor equally well independent of the relevance of the suggested commands to the subjects' immediate activities. She suggested that developers not concentrate on designing rules to make a computer assistant more intelligent. Rather, by allowing users to turn system proactivity off and on, developers could alleviate any negative reaction and still give the user chances to learn new things.

The results of our study suggest a number of follow-on experiments including issues of timing of agent advice. It would be interesting to degrade the quality of both the proactive advice given by the assistant and the reactions to queries posed by the user in order to see how this affects user perceptions. How much of a performance or recognition drop-off will begin to affect user performance or subjective assessment?

In this study, the interface assistant was an animated 3D representation without any programmed personality behavior. We could modify both the agent appearance and behavior to see if either affects user perceptions. Four of the participants, when asked their views of Clippy the Office assistant, tempered negative perceptions by saying that they liked the alternative cat or dog representation better, a comment we have heard anecdotally from colleagues as well. This reaction suggests that agent representation may provoke some subjective reaction.

Participants in our study were university undergraduates, a technologically savvy user population. Utilizing other participant demographics such as novice computer users or elderly individuals may generate different results.

6 Conclusion

This research examined the effects of interface assistants as compared to traditional documentation in helping people use a software tool. We found that performance with both reactive and proactive assistants was equivalent to that with printed help. Proactive suggestions made by the assistant did not

improve performance, but were viewed as being helpful by study participants.

Interface agents, as illustrated here, show promise but the space of factors that can influence performance and perception is large (Catrambone et al., 2002). We intend to continue to explore key features of agents, users, and tasks in hopes of developing design principles for interface agents.

Acknowledgments

References

Cassell, J. (2000), Embodied conversational interface agents. *Communications of ACM 43*, 4, 70-78.

Cassell, J., Bickmore, T., Billinghurst, M., Campbell, L., Chang, K., Vilhjálmsson, H., & Yan, H. (1999), Embodiment in Conversational Interfaces: Rea, *Proceedings of CHI '99*, 520-527.

Catrambone, R., Stasko J., & Xiao, J. (2002), Anthropomorphic Agents as a User Interface Paradigm: Experimental Findings and a Framework for Research, *Proceedings of the 24th Annual Conference of the Cognitive Science Society*, 166-171.

Dahlback, N., Jonsson, A., & Ahrenberg, L. (1993), Wizard of Oz Studies – Why and how, *Proceedings of the 1993 International Workshop on Intelligent User Interfaces*, 193-200.

Dehn, D.M. & van Mulken, S. (1999), The impact of animated interface agents: A review of empirical research. *International Journal of Human-Computer Studies 52*, 1, 1-22.

Fischer, G., Lemke, A., & Schwab, T. (1985), Knowledge-based help systems, *Proceedings of CHI '85*, 161-167.

Horvitz, E (1999), Principles of Mixed-Initiative User Interfaces, *Proceedings of CHI '99*, 159-166.

Isbister, K., & Doyle, P. (2002), Design and Evaluation of Embodied Conversational Agents: A Proposed Taxonomy, *AAMAS Workshop: Embodied Conversational Agents – Let's Specify and Evaluate Them!*

Lanier, J. (1995), Agents of alienation. *Interactions 2*, 3, 66–72.

Laurel, B. (1990), Interface agents: Metaphors with character, *The Art of Human-Computer Interface Design*, Laurel, B. (ed.), Addison-Wesley, 355-365.

Lester, J., Converse, S., Kahler, S., Barlow, T., Stone, B., & Bhogal, R. (1997), The Persona Effect: Affective Impact of Animated Pedagogical Agents, *Proceedings of CHI '97*, 359-366.

Lieberman, H. (1997), Autonomous interface agents, *Proceedings of CHI '97*, 67-74.

Mackay, W.E. (2001), Does Tutoring Really Have to be Intelligent? CHI '01 Extended Abstracts, 265-266.

Mayer, R., & Anderson, R. (1992), The instructive animation: Helping students build connections between words and pictures in multimedia learning. *Journal of Educational Psychology 84*, 4, 444-452.

McCrae, R., & Costa, P. (1987), Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology 52*, 1, 81–90.

Mousavi, S., Low, R., & Sweller, J. (1995), Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology 87*, 2, 319-333.

Nass, C., Steuer, J., & Tauber, E. (1994), Computers are social actors, *Proceedings of CHI '94*, 72-78.

Rickenberg, R. & Reeves, B. (2000), The effects of animated characters on anxiety, task performance, and evaluations of user interfaces, *Proceedings of CHI '00*, 329-336.

Rhodes, B. (2000), Margin Notes: Building a Contextually Aware Associative Memory, *Proceedings of IUI'00*, 219-224.

Ruttkay, Z., Dormann, C., & Noot, H. (2002), Evaluating ECAs – What and how? *AAMAS* Workshop *Embodied Conversational Agents – Let's Specify and Evaluate Them!*

Sanders, G., & Scholtz, J. (2000), Measurement and Evaluation of Conversational Agents, In *Embodied Conversational Agents*, Cassell, J., Sullivan, J., Prevost, S. & Churchill, E. (eds.), MIT Press, Cambridge MA, 346-373.

Schaumburg, H. (2001), Computers as tools or as social actors?- The users' perspective on anthropomorphic agents. *International Journal on Cooperative Information Systems* 10 (1,2), 217-234.

Selker, T. (1994), Coach: A Teaching Agent that Learns. Communications of the ACM 37, 7, 93-99.

Shneiderman, B (1997), Direct Manipulation Versus Agents: Paths to Predictable, Controllable, and Comprehensible Interfaces, In *Software Agents*, Bradshaw, J. (ed.), MIT Press, Cambridge MA, 97-106.