APPLIED COGNITIVE PSYCHOLOGY
Appl. Cognit. Psychol. (2008)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/acp.1527

Embodied Conversational Agents as Conversational Partners

MAX M. LOUWERSE[1]*, ARTHUR C. GRAESSER[1], DANIELLE S. McNAMARA[1] and SHULAN LU[2]

[1]*University of Memphis, USA*
[2]*Texas A & M University-Commerce, USA*

SUMMARY

Conversational agents are becoming more widespread in computer technologies but there has been little research in how humans interact with them. Two eye tracking studies investigated how humans distribute eye gaze towards conversational agents in complex tutoring systems. In Study 1, participants interacted with the single-agent tutoring system AutoTutor. Fixation times showed that the agent received most attention throughout the interaction, even when display size was statistically controlled. In Study 2, participants interacted with iSTART. Fixations were on the relevant agents when these agents spoke. Both studies provided evidence that humans regard animated conversational agents as conversational partners in the communication process. Copyright © 2008 John Wiley & Sons, Ltd.

Embodied conversational agents (ECAs) are animated characters that emulate human behaviour and communication. There has been a boom of these agents in computer environments that vary from conventional web systems to advanced virtual worlds. A fundamental question arises as to how humans interpret these agents. Are these ECAs distracting artefacts or human-like conversational partners? Answers to these questions require an in-depth analysis of the humans' perceptions, attention and interactions with the agents. The present studies were designed to address these questions.

Some studies have suggested that users interact with ECAs as they would with humans (Reeves & Nass, 1996), presumably because these computer-generated characters demonstrate many of the same multi-modal properties as do humans in face-to-face conversation (Cassell, Sullivan, Prevost, & Churchill, 2000). Studies in cognitive psychology, social psychology, artificial intelligence and education have investigated the effects of ECAs on user impressions, comprehension and learning gains (Atkinson, 2002; Baylor & Ryu, 2003; Graesser, Moreno, Marineau, Adcock, Olney, & Person, 2003). Some studies have shown that there are conditions in which users interacting with ECAs displayed on a computer monitor benefit from these interfaces more than from interfaces without ECAs. André, Rist, and Müller (1998) reported that users considered ECAs to be more helpful and entertaining than systems without ECAs. Lester, Converse, Stone,

*Correspondence to: Max M. Louwerse, Department of Psychology, Institute for Intelligent Systems, University of Memphis, 202 Psychology Building, Memphis, TN 38152-3230, USA. E-mail: mlouwers@memphis.edu

Kahler, and Barlow (1997) reported that ECAs can enhance problem solving skills in middle school children; Moreno, Mayer, Spires, and Lester (2001) found that students communicating with ECAs performed better and showed higher levels of motivation and interest than did those in a comparable text-only condition. Atkinson (2002) reported that ECAs improve learning when compared to text-only and voice-only conditions. Lusk and Atkinson (2007) replicated this latter finding by showing that embodied agents fostered learning compared to a voice-only condition. Other studies have come to the same conclusion, showing that ECAs help learning over printed text and speech alone (Graesser, Moreno, et al., 2003).

Nevertheless, other studies have presented a less conclusive picture. Modalities other than the ECA itself may foster learning (Craig, Gholson, & Driscoll, 2002). Graesser, Ventura, Jackson, Mueller, Hu, and Person (2003) reported no effect of conversational agents over print alone or speech alone in the context of navigational guides on the use of a complex learning environment on research ethics. Dehn and Van Mulken (2000) argued that the agent's impact on user performance and engagement is somewhat inconclusive, whereas Shneiderman (1997) and Shneiderman and Plaisant (2004) went so far as to conclude that an ECA is substantially less efficient than direct interactions with a computer. According to these alternative viewpoints, it is a debatable question whether the presence of ECAs facilitates comprehension and learning.

The role of agents in the human–computer interaction is not entirely clear within and across studies. Although Lusk and Atkinson (2007) reported that ECAs increase learning more than a voice-only condition, they did not find differences between fully (dynamic) and minimally (static) ECAs. In a similar vein, Wagenaar, Schreuder, and van der Heijden (1985) compared the delivery of weather forecasts by an ECA on TV and voice on radio and found no memory differences in favour of TV, which included entertaining graphics, whereas the radio did not. These findings suggest that sheer attention to humans or computer agents throughout an interaction does not present an adequate picture of its psychological impact. Pezdek and Hartman (1983) reported that 5-year-old pay attention to a television screen only when necessary and very much rely on auditory cues to determine when to pay attention to the TV screen. Thus, there is a need for studies to investigate at what points of time participants interact with the agent. Such studies will provide clues regarding the potential benefits that ECAs add to dynamic interactions.

The benefits of ECAs are also debatable according to cognitive load theory (Sweller, Van Merriënboer, & Paas, 1998). On the one hand, agents might add to an extraneous cognitive load because the user needs to process multiple sources of information (resulting in a potential split attention effect) and these sources also provide very similar information (face, voice, sometimes text), which may result in a redundancy effect. On the other hand, a well-designed ECA may reduce cognitive load, because it guides the learner regarding what to pay attention to and also provides multiple modalities, which may reinforce each other, resulting in a modality effect.

A fundamental question, and one that serves as a prerequisite to other questions related to processing, comprehension and learning, is how users pay attention to ECAs. Two alternative predictions can be made. The first prediction is motivated by an *information-only* hypothesis. This hypothesis states that the ECA will not attract attention or will lose attracted attention after an initial novelty phase because it does not provide additional information to the communication. The extreme perspective of this prediction is supported by studies showing that the spoken verbal modality carries most of the information in typical communication tasks, compared to the face which carries information primarily

DOI: 10.1002/acp

concerning emotions (Fish, Kraut, Root, & Rice, 1993; Reid, 1977). A less extreme perspective is that participants pay attention to humans and agents, but they are not a persistent attention magnet (Lusk & Atkinson, 2007; *cf.* Pezdek & Hartman, 1983).

The second prediction is what we refer to as the *conversational-partner* hypothesis, which has an analogue to everyday interactions with people. In face-to-face conversations the speaker's mouth shape and eyes attract the interlocutor's eye gaze (Argyle & Cook, 1976; Massaro & Cohen, 1983; Summerfield, 1987) for various communicative purposes (Clark, 1996). Users look for perceptual cues (pointing, gestures) to guide their attention and references to entities and events in the external world. According to this hypothesis, ECAs attract attention and do this over the course of a conversation. An agent provides a coherent centre of cognitive activities that persists throughout the learning experience.

There is little information on what human dialogue partners look at in face-to-face conversations. Early studies suggested that the face attracts attention in face-to-face dialogue (Argyle & Cook, 1976; Kendon, 1980), but there have been few online studies testing this hypothesis. Gullberg (2003) reported eye tracking evidence that the face of the dialogue partner dominates as a target of visual attention, supporting the conversational-partner hypothesis in the context of human–human communication.

The current study tested the information-only and conversational-partner hypotheses in two eye tracking studies. In Study 1, participants interacted with the intelligent conversational tutoring system AutoTutor (Graesser et al., 2004) that uses an ECA. In Study 2, participants interacted with the Interactive Strategy Training for Active Reading and Thinking (iSTART; McNamara, Levinstein, & Boonthum, 2004), which uses multiple ECAs. AutoTutor and iSTART both are intelligent tutoring systems that have conversations with students, both use ECAs, and both have shown to yield learning gains for thousands of students (Graesser, McNamara, & Van Lehn, 2005).

STUDY 1

AutoTutor is an intelligent tutoring system that assists students in actively constructing knowledge by holding a conversation in natural language. The system poses questions or problems that require approximately a paragraph of information from a student as an answer. A conversation between AutoTutor and the learner typically lasts 30–100 turns while solving a single problem.

There are four components, or sources of information, in the interface of AutoTutor, which is illustrated in Figure 1.

(1) An ECA is presented on the left region of the screen. The agent has a male talking head, with the image extending from the shoulders to above the head and wide enough to pick up most hand gestures.
(2) A main deep reasoning question is presented in the wide horizontal window at the top of the screen, right above the agent. This general question remains at the top of the web page until it is answered in a multi-turn dialogue between tutor and student. Answers to this main question typically involve 30–100 conversational turns between AutoTutor and the student.
(3) The student contributions are typed at the bottom wide horizontal window of the screen. Student contributions can consist of a variety of speech acts, including

DOI: 10.1002/acp

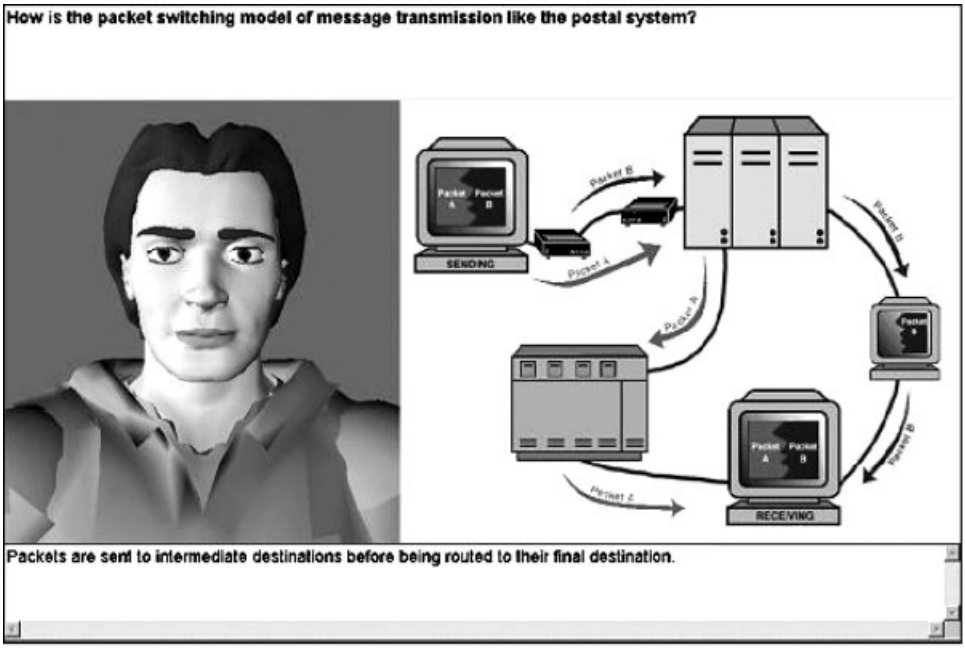

Figure 1. AutoTutor interface with four information sources: ECA, question, student input and graphic display

assertions and questions. As the student types in information for a conversational turn, the words are printed in a wide window at the bottom of the computer display.

(4) A graphic display shows components of the computer system that are relevant to the main question. The position of this window is in the right region of the computer display, next to the ECA. Graphic displays are associated with approximately a third of the questions. For the remaining two-thirds, this window remains blank.

According to the conversational-partner hypothesis, participants primarily look at the ECA because it serves as a participant in the dialogue. Alternatively, according to the information-only hypothesis, looking at the agent is not necessary because the participants can listen to the ECAs speech. An eye tracking study was conducted to measure the fixation times on AutoTutor's four information sources as a function of four 5-minute time intervals in a 20-minute tutoring session.

Method

Participants

Twelve undergraduate students at the University of Memphis participated for extra credit. Participants had normal or corrected-to-normal vision.

Materials

Each participant had an interactive tutorial session with AutoTutor on computer literacy (Figure 1). At the start of a session, AutoTutor introduced itself and explained how to

DOI: 10.1002/acp

interact with the system. It then started the tutoring session with a deep reasoning question. The student worked towards answering the question through interactions with the system in natural dialogue. The ECA was present on the screen at all times, as was the question and the text box where students could type their input. For some subtopics, graphic displays were used in the session for illustrative purposes.

Apparatus

During the interaction with AutoTutor, the participant's eye movements were tracked using an ASL (Applied Science Laboratory) Model 501 eye tracker over a 20-minute time span. This eye tracker had a temporal resolution of 60 Hz. The light head mounted optics recorded the eye whereby the centre of the pupil and the corneal reflection were tracked to determine the relative position of the eye. A magnetic head tracking equipment (Ascension Flock of Birds) was used in order to compensate for possible head movements. Accuracy of the gaze position record is about 0.5 degrees visual angle. Participants were calibrated using a 9-point grid both before the session began and throughout the session to ensure reliable data. They were seated about 700 mm in front of the stimulus monitor.

Results and discussion

Our first analysis measured the total fixation time on the four information sources of the display as a function of the four 5-minute time intervals. The raw time scores were converted to proportions. That is, we computed the proportion of time that the eyes fixated on the ECA, the question, graphic display, student input and the keyboard. Figure 2 shows the proportion of fixation times as a function of the five areas of interest and the four 5-minute intervals. An analysis of variance (ANOVA) was performed on these proportions, using an information source (agent, question, answer, display) by time-interval factorial design; the off-screen times were not included in the analysis in order to remove the degrees of freedom problem when proportions add to 100%. There was a significant main effect of information source ($F(3,33) = 31.11$, MSE = 1514.70, $p < .01$, $\eta_p^2 = .74$), but no significant main effect of time interval, and no significant interaction between information source and time interval.

The previous analysis did not control for the size of the windows. Indeed, the display of the agent is larger (28% of display) than the student input box or question box (both 13% of display), but smaller than the graphic display (46% of display). To determine whether the student's attention to a window is more than would be expected by chance (i.e. if eye

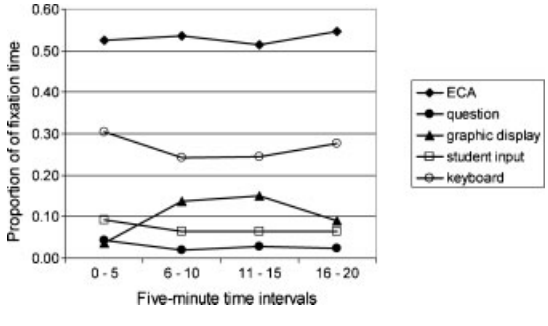

Figure 2. Allocation of eye fixations to different regions of the display (in proportions)

DOI: 10.1002/acp

movements were randomly allocated to regions on the display), we computed adjusted fixation ratios. For example, if the agent window took up 28% of the real estate on the display, but 56% of fixation time was on the agent (when considering the four windows on the computer screen), we would conclude that the agent attracted the attention of the participant above chance level (adjusted fixation ratio = .56/.28 = 2.0). The adjusted fixation ratio is 1 if the eyes wandered randomly within the computer screen, >1 to the extent that an agent is a conversational partner and <1 to the extent the agent is ignored.

Adjusted fixation ratios were 1.81 for the ECA, .43 for the question, .68 for the student input display and .78 for the graphic display. These results again revealed a visual attention effect for the agent ($F(3,33) = 15.78$, MSE = 738.33, $p < .01$, $\eta_p^2 = .59$), supporting the conversational-partner hypothesis.

The fixation ratios and adjusted fixation ratios showed that the ECA serves as a conversational partner. This effect cannot simply be attributed to novelty because the ECA remained on the monitor throughout the tutoring session and the distribution of attentional resources did not vary over time. Although graphic displays presented on the screen obviously attract some of the user's visual attention, the user negotiated this attention between agent and display. This negotiation is similar to the user negotiating fixations between face and gesture in face-to-face interactions, whereby the face dominates visual attention (Gullberg, 2003). The reason for face dominance in visual attention is the hearer indicating attention and interest to the speaker (Argyle & Cook, 1976; Kendon, 1980).

The AutoTutor agent is an animated figure rather than a static figure. It may be the case that just the animated parts of face-like figure attract fixation. In that case, the ECA attracts attention, as predicted by the conversational-partner hypothesis, but not because of face dominance in visual attention, but simply because of animation. In AutoTutor's ECA, the mouth is the most animated part of the face. If the moving part of the screen attracts attention, participants' fixations should concentrate on the mouth. Alternatively, if participants looked at AutoTutor's face as if they were looking for perceptual cues in a human face, fixations would primarily centre on the nose bridge of the speaker's face capturing the eyes and mouth of the speaker simultaneously (Gullberg, 2003). The face of AutoTutor was divided into five areas of interest: its two eyes, nose, mouth, forehead, both cheeks and its shoulders (see Figure 1). Accounting for the size of the areas, fixations primarily took place on the (non-animated) nose bridge compared to the other areas ($F(4, 44) = 6.49$, MSE = 154.33, $p < .001$, $\eta_p^2 = .37$), as shown in Table 1.

These findings suggest that the speaker's face is perceived in the same way as the face of a conversational partner in human face-to-face conversation; or at the very least, it is a close analogue. Consequently, we can predict that in scenarios with multiple ECAs, eye gaze will be on those ECAs that are relevant to the conversation. This hypothesis was tested in Study 2.

STUDY 2

Study 2 used a multi-agent interface and tested whether participants pay attention to the relevant agent at the relevant times. If users interact with ECAs as they would with other humans, they would fixate on the agent only while that agent is speaking or is referred to. Participants interacted with the intelligent tutoring system iSTART. This system provides training to help students more effectively self-explain difficult texts while reading. iSTART

DOI: 10.1002/acp

Table 1. Proportion fixation times and adjusted fixation ratios to agent

	Proportion fixation times	Adjusted fixation ratios
Eyes	.27	9.43
Nose	.24	24.12
Mouth	.03	2.24
Shoulders	.19	1.46
Cheeks	.14	7.53
Shoulders	.14	1.46

Note: Adjusted fixation ratios = screen real estate divided by fixation times.

delivers reading strategy training using an interactive and adaptive format. In the introduction phase of the system, three ECAs interact with each other and with the student to increase active processing and participation (McNamara et al., 2004).

Two components in the iSTART interface are relevant for the current study, as illustrated in Figure 3.

(1) Three ECAs are presented on the screen, the main character (instructor) on the left centre of the screen and two additional characters (students) on the right centre.
(2) A text balloon appears above the agent as soon as the agent starts speaking. These text balloons contain the transcriptions of the spoken text in a readable font size (Arial 11).

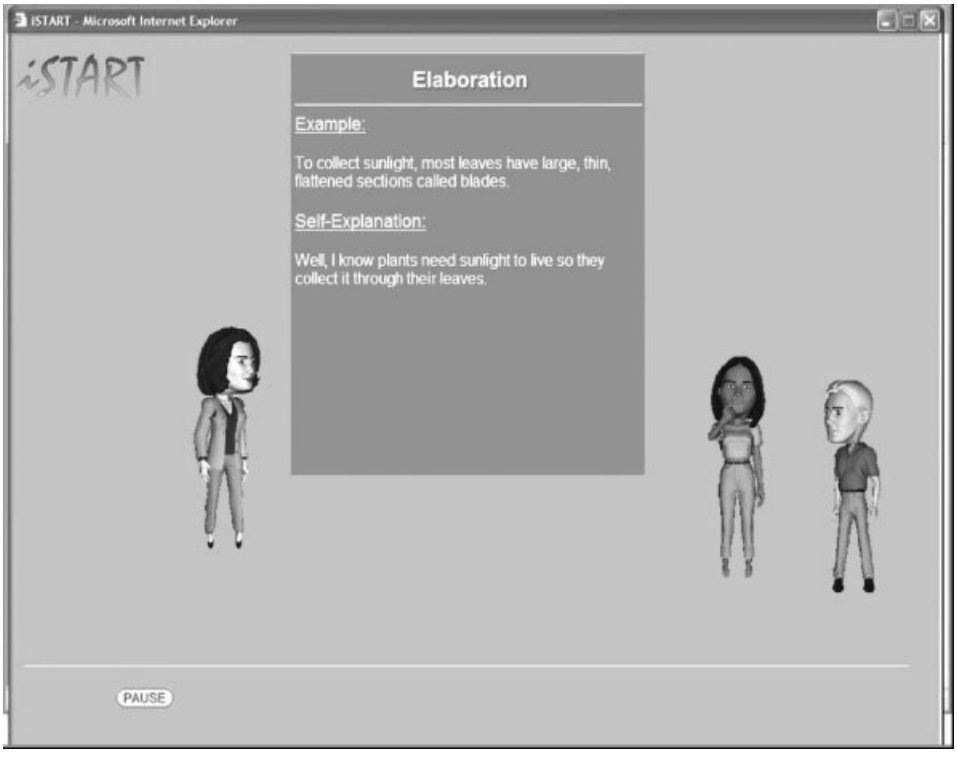

Figure 3. iSTART interface with three ECAs

DOI: 10.1002/acp

The system has more than these four components, but Study 2 focussed on a comparison of the iSTART agents only. The conversations consisted of seven phases, each covering a particular reading strategy: (1) overview; (2) self-explanation; (3) monitoring; (4) paraphrase; (5) prediction; (6) elaboration and (7) bridging. For the purposes of the current study, details concerning the specific reading strategies are not relevant (see McNamara et al., 2004, for detailed information). The chronology of these phases is of importance, however, because it allows us to identify whether fixation times on the areas of interest increase, decrease or remain the same over time.

Participants interacting with iSTART can look at the three ECAs, at the text balloon or at other areas on the screen at any point in time. Viewing the agents is actually not needed because the participant can listen to the speech and can even read the transcribed speech in the text balloon. According to the conversational-partner hypothesis, the participant will look at the ECA only when the ECA speaks. In cases where participants look at the text balloon, they will negotiate their attention between the information in the balloon and the ECA.

Study 2 investigated whether the participants look at the relevant parts of the screen at the relevant time. Instead of conducting analyses on differences between the fixation times, we opted for a precision and recall analysis on the number of times the participant looked at the ECA when this was relevant for the conversation. High precision and recall would suggest that participants pay attention to the relevant agent at the relevant time, as predicted by the conversational-partner hypothesis. A low precision and low recall suggests that participants do not look at the relevant agent at the relevant time, indicating evidence for the information-only hypothesis.

Method

Participants

Seven undergraduate students at the University of Memphis participated in this study and received extra credit in an undergraduate psychology course. Participants had normal or corrected-to-normal vision.

Materials

The participants interacted with iSTART during its introduction phase. The participants clicked on a button at the bottom of the screen to move from one screen to the next.

Apparatus

During the interaction with iSTART, the participant's eye movements were recorded using an SMI iView X Hi-Speed eye tracker with a temporal resolution of 240 Hz. This system consists of a tracking column, which contains the camera and an infrared light source, and a chin rest to keep the participant's head still. Accuracy of the gaze position recording is approximately a 0.25–0.50 degrees visual angle. The system has the advantage of recording a video of the stimulus (iSTART interaction) with a superimposed eye tracking overlay. Participants were calibrated using a nine-point grid both before and throughout the session to ensure reliable data. The computer monitor was placed about 500 mm in front of the subject.

DOI: 10.1002/acp

Results and discussion

Participants spent an average of 29.5 minutes (SD = 1.35) interacting with iSTART while their eye movements were recorded. These data were then coded and analysed by two Psychology research assistants at the University of Memphis who performed a precision and recall analysis (1) when the participant looked at the target when the target was relevant, (2) when the participant did not look at the target when the target was relevant and (3) when the participant looked at the target when the target was not relevant. This information was coded in 250 milliseconds intervals.

Interrater reliability ($N = 2925$) was near perfect (Cohen's $\kappa = .96$) between the two human judges. Kappas did not differ between the stages of the introduction (.94–.98) or between the characters (.96–.98).

Precision, recall and *F*-scores are presented in Table 2. Precision is the number of times participants looked at the relevant ECA when they were supposed to look at that ECA divided by the number of times they looked at all ECAs. Recall is the eye gaze on the relevant ECA divided by the number of times they should have looked at that ECA. That is, if participants solely looked at one ECA, precision for that one ECA would be high, but recall (i.e. coverage of all ECAs the person is supposed to look at) low. At the same time, if participants continuously scanned the page looking at all ECAs, recall would be high, but precision low (i.e. coverage is high, at the cost of looking at the relevant ECA at the right time). The *F*-score is the weighted harmonic mean of precision and recall (Baeza-Yates & Ribeiro-Neto, 1999).

The precision and recall scores suggested that participants looked at the ECAs and the corresponding text balloons at the relevant times. These results show that participants negotiated their eye gaze between the ECAs and text balloons. No significant differences were found between the three characters or the text balloons (all Fs $< .25$), suggesting that participants interacted with the three ECAs in similar ways. Moreover, no significant differences were found between the seven reading strategy stages (all Fs $< .5$), suggesting that participants looked at the agents in similar ways over time.

ECAs sometimes referred to each other during the iSTART session. For instance, the instructor ECA introduced the students ECAs. The text balloon obviously only accompanies the speaker and not the ECAs introduced, so there is no information-based reason to look away from the text balloon of the speaker agent, as predicted by the information-only hypothesis. Nevertheless, precision, recall and *F*-scores for participants looking at the ECAs when these were referred to were high (.72, .89, .78, respectively), suggesting that participants look at the ECAs whenever this is relevant for the conversation.

These results support the conclusion from Study 1 that participants look at ECAs because these serve as conversational partners. This is not different for single or multiple

Table 2. Precision, recall and *F*-score on iSTART targets

	ECA as target			Text balloon as target			
	Instructor	Student 1	Student 2	Instructor	Student 1	Student 2	Total
Precision	1.00	.93	.95	1.00	.93	.98	.96
Recall	.66	.72	.66	.91	.93	.93	.80
F-score	.81	.67	.68	.97	.81	.90	.81

Note: Recall for ECA and text balloon can exceed 100% because quick fixations between them can co-occur within the same coding time frame.

DOI: 10.1002/acp

ECAs. The focus on the speaker agent occurs when the ECA is speaking or is referred to, and this attention does not wane over time.

GENERAL DISCUSSION

Many studies have investigated the effect of ECAs on comprehension, likeability and learning, some showing benefits and some showing drawbacks. There are a number of possible reasons why the jury is still out. The agent's character, its gender, its voice and the content, could all contribute to inclusive findings (Louwerse, Graesser, Lu, & Mitchell, 2005). The inconclusiveness of the findings could also be explained by a user's behaviour with regard to the conversational dimensions of the agents. Surprisingly, no research has investigated how users allocate their attention to the agents during the course of interacting with these agent-based learning environments.

Our findings show that ECAs have the effect of a conversational partner, and that this effect does not wear off over time with participants spending a substantial proportion of time looking at the agent. Connecting this to the inconclusive effects on learning outcomes found in other studies, we might speculate that potential positive or negative effects on learning are related to the quality of information provided by the agent. The agent will attract attention when speaking; in other words, it draws attention away from other information sources. According to cognitive load theory (Sweller et al., 1998), this will only foster learning when the information provided by the agent at that particular moment fosters learning (*cf.* germane load), and may hamper learning when it is in any way redundant or not helpful for learning (extraneous load). Our study did show though that negative effects on learning due to the mere presence of (multiple) agents is unlikely, as learners pay very little attention to them when they are not speaking or are not being referred to.

This paper has equated eye gaze and visual attention. The question could be raised whether the current results support overt (top-down processing) or covert (bottom-up processing) attention (Hunt & Kingstone, 2003; Wright & Ward, 2008). This question falls outside the scope of this paper, but the data appear to support overt visual attention. In Study 1, covert visual attention would have directed the participant's eye gaze continuously to AutoTutor, not the visual display, and to its moving mouth, not its static nose. In Study 2, covert visual attention would have moved participants' eye gaze to the dynamic text balloon, not to the static agent. Results show different patterns, indicating voluntary visual attention that is aligned with social interaction (see also Louwerse et al., 2005).

ECAs have become more common in a variety of human–computer interaction tasks. Though there is increasing evidence that students learn from interacting with these systems (VanLehn, Graesser, Jackson, Jordan, Olney, & Rose, 2007) and there is mounting evidence that the communicative modalities of these agents matter (Louwerse et al., 2005), it has been unknown whether we interact with these agents online as we do with humans in face-to-face conversations. The results of the present study demonstrate that ECAs guide participants' perceptual attention, just as they are guided by humans in conversational dialogue.

ACKNOWLEDGEMENTS

This research was supported by grants from National Science Foundation (IIS-0416128, ITR 0325428, REC 0106965, REC 0126265, REC-0089271) and IES (R305G040046). We thank Courtney Bell for her help in the data collection.

DOI: 10.1002/acp

REFERENCES

André, E., Rist, T., & Müller, J. (1998). Integrating reactive and scripted behaviors in a life-like presentation agent. In K. P. Sycara, & M. Wooldridge (Eds.), *Proceedings of the second international conference on autonomous agents* (pp. 261–268). Minneapolis: ACM Press.

Argyle, M., & Cook, M. (1976). *Gaze and mutual gaze*. Cambridge: Cambridge University Press.

Atkinson, R. K. (2002). Optimizing learning from examples using animated pedagogical agents. *Journal of Educational Psychology*, *94*, 416–427.

Baeza-Yates, R., & Ribeiro-Neto, B. (1999). *Modern information retrieval*. New York: Addison-Wesley.

Baylor, A. L., & Ryu, J. (2003). Does the presence of image and animation enhance pedagogical agent persona? *Journal of Educational Computing Research, 28*, 373–395.

Cassell, J., Sullivan, J., Prevost, S., & Churchill, E. (2000). *Embodied conversational agents*. Cambridge, MA: MIT Press.

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.

Craig, S. D., Gholson, B., & Driscoll, D. (2002). Animated pedagogical agents in multimedia educational environments: Effects of agent properties, picture features, and redundancy. *Journal of Educational Psychology*, *94*, 428–434.

Dehn, D. M., & Van Mulken, S. (2000). The impact of animated interface agents: a review of empirical research. *International Journal of Human-Computer Studies*, *52*, 1–22.

Fish, R. S., Kraut, R. E., Root, R. W., & Rice, R. (1993). Evaluating video as a technology for informal communication. *Communications of the ACM*, *36*, 48–61.

Graesser, A. C., Lu, S., Jackson, G. T., Mitchell, H. H., Ventura, M., Olney, A., et al. (2004). AutoTutor: A tutor with dialogue in natural language. *Behavior Research Methods, Instruments, and Computers*, *36*, 180–193.

Graesser, A. C., McNamara, D. S., & VanLehn, K. (2005). Scaffolding deep comprehension strategies through Point&Query, AutoTutor, and iSTART. *Educational Psychologist*, *40*, 225–234.

Graesser, A. C., Moreno, K., Marineau, J., Adcock, A., Olney, A., & Person, N. (2003). AutoTutor improves deep learning of computer literacy: Is it the dialog or the talking head? In U. Hoppe, F. Verdejo, & J. Kay (Eds.), *Proceedings of artificial intelligence in education* (pp. 47–54). Amsterdam: IOS Press.

Graesser, A. C. Ventura, M., Jackson, G. T., Mueller, J., Hu, X., & Person, N. (2003). The impact of conversational navigational guides on the learning, use, and perceptions of users of a web site. *Proceedings of the AAAI spring symposium 2003 on agent-mediated knowledge management* (pp. 9–14). Palo Alto, CA: AAAI Press.

Gullberg, M. (2003). Eye movements and gestures in human face-to-face interaction. In J. Hyona, R. Radach, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movements* (pp. 685–703). Oxford: Elsevier Science.

Hunt, A. R., & Kingstone, A. (2003). Covert and overt voluntary attention: Linked or independent? *Cognitive Brain Research*, *18*, 102–105.

Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In M. R. Key (Ed.), *The relationship of verbal and nonverbal communication* (pp. 207–227). The Hague: Mouton.

Lester, J., Converse, S., Stone, B., Kahler, S., & Barlow, T. (1997). Animated pedagogical agents and problem-solving effectiveness: A large-scale empirical evaluation. *Proceedings of the eighth world conference on artificial intelligence in education* (pp. 23–30). Amsterdam: IOS Press.

Louwerse, M. M., Graesser, A. C., Lu, S., & Mitchell, H. H. (2005). Social cues in animated conversational agents. *Applied Cognitive Psychology*, *19*, 1–12.

Lusk, M. M., & Atkinson, R. K. (2007). Varying a pedagogical agent's degree of embodiment under two visual search conditions. *Applied Cognitive Psychology*, *21*, 747–764.

Massaro, D. W., & Cohen, M. M. (1983). Integration of visual and auditory information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *9*, 753–771.

McNamara, D. S., Levinstein, I. B., & Boonthum, C. (2004). iSTART: Interactive strategy trainer for active reading and thinking. *Behavioral Research Methods, Instruments, and Computers*, *36*, 222–233.

DOI: 10.1002/acp

Moreno, R., Mayer, R. E., Spires, H. A., & Lester, J. (2001). The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction*, *19*, 117–213.

Pezdek, K., & Hartman, E. F. (1983). Children's television viewing: Attention and comprehension of auditory versus visual information. *Child Development*, *54*, 1015–1023.

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge: Cambridge University Press.

Reid, A. (1977). Comparing telephone with face-to-face contact. In I. de Sola Pool (Ed.), *The social impact of the telephone* (pp. 386–415). Cambridge, MA: MIT Press.

Shneiderman, B. (1997). Direct manipulation versus agents: Paths to predictable, controllable and comprehensible interfaces. In J. M. Bradshaw (Ed.), *Software agents* (pp. 97–106). Menlo Park, CA: AAAI Press.

Shneiderman, B., & Plaisant, C. (2004). *Designing the user interface: Strategies for effective human-computer interaction*. Boston: Addison-Wesley.

Summerfield, A. (1987). Some preliminaries to a comprehensive account of audio-visual speech perception. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 3–52). Hillsdale, NJ: Erlbaum.

Sweller, J., Van Merriënboer, J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*, 251–296.

VanLehn, K., Graesser, A. C., Jackson, G. T., Jordan, P., Olney, A., & Rose, C. P. (2007). When are tutorial dialogues more effective than reading? *Cognitive Science*, *30*, 3–62.

Wagenaar, W. A., Schreuder, R., & van der Heijden, A. H. (1985). Do TV pictures help people to remember the weather forecast? *Ergonomics*, *28*, 765–772.

Wright, R. D., & Ward, L. M. (2008). *Orienting of attention*. Oxford: Oxford University Press.

DOI: 10.1002/acp