

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

Social Cues in Animated Conversational Agents

Max M. Louwerse

Arthur C. Graesser

Shulan Lu

Heather H. Mitchell

University of Memphis

Address for correspondence:

Max M. Louwerse

Department of Psychology /

Institute for Intelligent Systems

University of Memphis

202 Psychology Building

Memphis, TN 38152

Phone: (901) 678-2143

Fax: (901) 678-2579

Email: mlouwers@memphis.edu

Abstract

In human-computer interaction people often interpret the interaction with the computer as interactions with humans. The social agency theory suggests that social cues like the face and voice of the agent motivate this interpretation. In two off-line experiments in which comprehension scores and liking ratings were collected, we found that participants preferred natural agents with natural voices, as predicted by the social-cue hypothesis. Although female agents with male voices formed an exception, this was explained by a stereotype effect. These findings support the social-cue hypothesis and the social agency theory that human characteristics are applied in the perception of computational animated conversational agents.

Intelligent dialog systems have received considerable attention in fields such as artificial intelligence, computer science, cognitive science, and computational linguistics. Throughout the years, these systems have become progressively more sophisticated, with intelligent systems for train control (Allen et al., 1995), travel bookings (Pellom, Ward, & Pradhan, 2000), and personnel work (Franklin, 2001). A subset of these intelligent dialog systems are intelligent tutoring systems. These are computer-based instructional systems that continuously make inferences about a student's mastery of topics, facts, rules and mental models in order to dynamically adapt instruction and to deliver the right information at the right time. Various tutoring systems have been tested on humans and have proven to facilitate learning. For example, there are well-tested tutors of computer literacy (such as AutoTutor: Graesser, Person, Harter, & TRG, 2001; Graesser, et al., 2004), physics (such as Andes: Gertner & VanLehn, 2000; DIAGNOSER: Hunt & Minstrell, 1996; or AutoTutor: Graesser, VanLehn, Rosé, Jordan & Harter, 2001), electronics (such as SHERLOCK: Lesgold, Lajoie, Bunzo, & Egan, 1992), reading strategies (iSTART: McNamara, Levinstein, & Boonthum, 2004), and U.S. federal policies and regulations (HURAA: Graesser, Hu et al., 2002). The latest generation of intelligent tutoring systems adopts conversational interfaces (Graesser, et al., 2004; Graesser, VanLehn et al., 2001; Louwerse, Graesser, Olney & TRG, 2002). Students interacting with these systems are engaged in natural conversations communicating with the computer as they would with a human.

The naturalness of the conversation can be maximized by the use of animated conversational agents, because of the availability of both linguistic (semantics, syntax) and paralinguistic (pragmatic, sociological) features. These animated agents have anthropomorphic, automated, talking heads with facial features and gestures that are coordinated with text-to-

speech-engines. That is, the computer controls the movement of the head, eyes, eye brows and cheek bones, as well as movement of the mouth that is in synchrony with the speech (Cassell & Thorisson, 1999; Massaro & Cohen, 1994; Picard, 1997). Examples of these agents are Baldi (Massaro & Cohen, 1994), COSMO (Lester, Stone & Stelling, 1999), STEVE (Rickel & Johnson, 1999), Herman the Bug (Lester, Stone, Stelling, 1997) and AutoTutor (Graesser, Person, et al., 2001). The embodiment of these conversational agents has become an important topic in research on human-computer interaction (HCI), psycholinguistics, psychology and cognitive science (Cassell, Sullivan, Prevost & Churchill, 2000).

Social Agency in Human-Computer-Interaction

Relevant questions in the development of intelligent systems concern the benefits of the use of an animated conversational agent. Do users enjoy interacting with a computational agent? Do users interact with the agent as they would interact with a human? Do animated conversational tutoring agents yield pedagogical benefits?

Some studies have presented a pessimistic picture and negative answers to these questions. There have been speculations that animated agents limit, or even interfere with, efficient human-computer interaction. Shneiderman (1992, 1997) for instance describes a variety of anthropomorphic designs that have been rejected by customers. He argues that direct interaction with a computer is more beneficial than indirect interaction through an animated agent, particularly because animated conversational agents will never reach full human-level intelligence. A similar negative view can be found in Dehn and Van Mulken's (2000) review of 30 empirical studies that use animated conversational agents. Dehn and Van Mulken conclude that the impact of these agents on user performance and engagement with the system show that the benefit of agents is at best inconclusive.

Nonetheless, there is more optimistic research showing that people interpret human-computer interaction as human-to-human interaction as long as sufficient social cues are provided. This idea goes back to Reeves and Nass's (1996) media equation. Because of a human tendency to confuse what is real with what is perceived to be real, people automatically use social rules to guide their actions with these media. The social agency theory argues that people interpret computers as social partners. Consequently theories from social psychology can be applied to human-computer interaction according to the social agency theory. Moreno and Mayer (2001) discuss the consequences when social cues in HCI encourage the interpretation of a computer as a social partner. On the one hand, a social-cue hypothesis suggests that if social cues like facial expressions or human voices are present, people will comprehend computers better and will rate them more favorably. On the other hand, an interference hypothesis suggests that social cues may backfire. In that case, people will comprehend less from computers with social cues because these cues interfere in the HCI.

Various social cues can be distinguished, ranging from the presence of the agent and its character to the agent's gender and its voice. André, Rist, and Müller (1999) for instance found support for the social-cue hypothesis, showing that users in fact consider animated conversational agents as more helpful and entertaining than systems without those agents. Lester, Converse, Stone, Kahler and Barlow (1997) reported that animated pedagogical agents can enhance problem solving skills in middle school children. Moreno, Mayer, Spires, and Lester (2001) found that students communicating with an animated pedagogical agent with a human voice performed better and showed higher levels of motivation and interest than a comparable text-only condition. Mayer, Sobko and Mautone (2003) showed that human voices

resulted in better learning performance and more favorable ratings than computer synthesized voices and that standard accents had similar advantages over foreign accents.

But sometimes results do not provide strong support for the social-cue hypothesis, though they do not provide support for the interference hypothesis either. Atkinson (2002) compared text-only, voice-only and voice-and-agent conditions to distinguish between animation and voice, as well as to determine effects of synthesized speech. He found some evidence that the latter condition fostered learning from examples but warned that further research on animation and speech synthesis systems is needed. Whereas Atkinson (2002) did find effects for learning, Craig, Gholson, and Driscoll (2002) compared the presence versus absence of an animated agent and did not find a difference in learning gains. A follow-up experiment suggested that the voice rather than the animated agent provided some learning gains. Graesser, Jackson, et al. (2003) came to a similar conclusion when they found that effects of the medium (i.e., animated agent, speech, text) remain more subtle than the content that is communicated. At the same time, Moreno, Klettke, Nibbaragandla, Graesser and TRG (2002) found that there is no significant correlation between how much an agent is liked and how much is learned from an agent, even though agents varied significantly on these two dimensions. Baylor and Ryu (2003) pointed out that different studies investigating animated agents often compare different modalities: human voice versus agent voice, still images versus animations; human-like agents versus cartoon-like agents. This makes a cross-comparison problematic. It may very well be the case that different social cues encourage a social agency interpretation differently.

The present paper seeks to determine what social cues contribute to social agency? Does it matter whether the agent is human-like? What is the role of the agent's gender? What

is the role of the agent's voice? These questions are addressed in comparing human-like and cartoon-like agents (Experiment 1), the gender of the agent (Experiment 1 and 2) and the quality of the speech engine (Experiment 2). The social-cue hypothesis predicts that those conditions with most social cues, for which it is easiest to apply human characteristics to the agent, yield the highest scores.

Experiment 1

Experiment 1 investigated some of the social-cues that may attract the user's attention. Participants received a narration on the Roman Empire presented by one of eight animated conversational agents. Two were human-like agents and two were cartoon-like agents, each pair consisting of a male and a female agent with either a male or female voice. Three agents were selected from the Microsoft Agent Ring characters (<http://www.msagentring.org/chars.htm>) whereas one agent was based on the Agent Ring Characters but generated by our research team for use in an intelligent system (see Graesser, et al., 2004). The agents included Charlie, a human-like female agent; Marco, a human-like male agent; Marge, a cartoon-like female agent, and Milton, a cartoon-like male agent. Pictures of these agents are presented in Figure 1. For each agent we used two voices, one male and one female voice, generated by the AT&T Labs Natural Voices Text-To-Speech System (<http://www.naturalvoices.att.com>). The combination of these 2 x 2 x 2 factors allowed us to examine the effects of type of agent (human versus cartoon), gender of agent (male versus female), gender of speech (male or female), and their interactions. The social-cue hypothesis

 FIGURE 1 ABOUT HERE

predicts that the computer conversations that are most similar to human conversations are comprehended best and rated most favorably: The human-like agents resemble the human conversations more than the cartoon-like agents, as do the gender matching voices compared to the gender-mismatching voices.

Method

Participants. Forty-eight undergraduate students, 20 males and 28 females, at the University of Memphis participated in this experiment for course credit.

Materials. Eight different videos were created with an animated conversational agent presenting various topics about the Ancient Roman Empire. Each video lasted approximately 4 minutes and was followed by ten multiple-choice comprehension questions and ten multiple-choice impression questions. Comprehension questions tapped into various aspects of the content that was presented, including local and global content. An example of a passage and a comprehension question are given below.

The reign of Augustus brought about a period of peace after many years of civil war and slave uprisings. These wars drained the resources of the wealthy, which were given the task of supporting huge armies outside their home borders, as well as growing agriculture into a decline. As Augustus took power, this general decrease in wealth was rapidly turned. A prosperous upper class emerged once again [...].

Why was the upper class financially strained by war?

- a. Because they couldn't sell their goods.
- b. Because they had to go to war themselves.
- c. Because they lost their workers.
- d. Because they had to support their armies.

The ten comprehension questions were followed by ten impression questions. As with the comprehension questions, each impression question used four multiple-choice answers. Two impression questions were created for the perception of language (e.g., how would you assess the agent's use of language?), and the inter-item reliability assessed alpha at .79. Similarly, a pair of questions was created on the perception how life-like the agent was (e.g., compared to a character in a computer game, how life-like is the agent?); the alpha reliability assessment was .94. The interaction with the agent was measured by two questions (e.g., how do you rate the interaction with the agent?), with an inter-item reliability of alpha = .94. Two questions were used to determine the expressiveness of the agent (e.g., how would you rate the agent's facial expressions?), and the inter-rater reliability assessed alpha at .95. Questions were also asked with regard to the entertainment value of the agent (e.g., how entertaining did you find the agent?), with an alpha of .99.

Procedure. The different agents were compared by having participants watch and listen to eight mini-lessons on the Roman Empire. The assignment of mini-lessons to participants and conditions was counter-balanced using an 8 x 8 Latin square design. Each presentation was comprised of one agent from the eight agent combinations, i.e., (human – cartoon) x (male – female agent) x (male voice – female voice), so that each participant was presented with all agents. Thus, each participant saw a human male agent with a male voice, a human male agent with a female voice, a cartoon male agent with a male voice, etc., with agent and mini-lesson counterbalanced across the eight conditions, to avoid content or order effects. To ensure full attention to the agent and the speech, each agent appeared on a black screen, and spoke for approximately three to four minutes through the participant's headphones.

The presentation began with an instruction, followed by an agent presenting a mini-lesson like the one presented above. After each video, participants completed a ten question multiple-choice test, quizzing them over material just covered by the agent. This test was followed by a multiple-choice questionnaire on the participant's impressions of the agent and the agent's quality of style and speech.

Results

A repeated measures analysis of variance (ANOVA), using the character (human/cartoon) x gender (male/female) x voice (male/female) factorial design, was conducted on the proportions of correct answers for the comprehension questions. Performance of the comprehension questions did not yield any significant main effects, but there was a significant three-way interaction between agent type, agent gender and voice gender, $F(1, 47) = 4.54, p < .05, MSE = 132.20, \eta^2 = .09$. Means and standard deviations are presented in Table 1. As predicted by the social-cue hypothesis the voices that match the gender of the agent yield higher comprehension scores than the mismatching voices, but contrary to the hypothesis this was only true for the cartoon-like agents. In fact, for the female human-like agent with the mismatching male voice scores were obtained that were equally high. Perhaps these effects can be explained by participants liking these agents more, as the impression scores could explain.

TABLE 1 ABOUT HERE

For the impression questions, we analyzed the data separately for each of the five groups: language use, life-likeness, interaction, expressiveness and entertainment. Not

surprisingly, life-like ratings were higher for human agents than cartoon agents ($M = 3.14$, $SD = 0.29$ vs. $M = 3.01$, $SD = 0.27$), $F(1, 47) = 7.39$, $p < .01$, $MSE = .28$. Ratings for life-likeness, interaction and expressiveness showed interactions for agent gender and voice gender ($F(1, 47) = 8.82$, $p < .01$, $MSE = .28$; $F(1, 47) = 3.70$, $p = .06$, $MSE = 1.81$; $F(1, 47) = 9.78$, $p < .01$, $MSE = .37$, respectively). The voice matching gender conditions received higher scores than the conditions when the voice did not match the gender. Impression scores showed the patterns expected by the social-cue hypothesis, with human agents (human-like) and voice matching the agent's gender (life-likeness, interaction and expressiveness) being rated more favorably. However, the advantage for the mismatching voice in human female agent can not be explained by the social-cue hypothesis.

Discussion

The results in Experiment 1 provided some evidence for the social-cue hypothesis. Comprehension scores were higher for the cartoon-agents where the voices matched their gender than for the mismatching voices. For the human-like agent such an effect was not obtained. On the contrary, the female agent with the male voice yielded higher comprehension scores. One explanation for this finding is that participants may have preferred the cartoon-like agents over the human-like agents, and the female agent with the male voice over the other three conditions. Impression ratings however cannot readily account for this. Human agents were rated more humans-like, and matching voices were rated most favorably.

We have found that social cues like the type of agent and whether the voice matches the agent's gender have an impact on comprehension scores and impression ratings, but patterns are not entirely in the expected directions. The interaction effect between agent gender and voice gender found for cartoon-like agents only was not expected. The fact that

cartoon-agents yielded higher scores may be explained by their posture. The availability of animated agents did not allow for human agents other than head-only human-like agents, whereas no cartoon-like agents were available that are not embodied. The strength of these embodied cartoon-like agents is that participants can rely on the social cue of gesture as the primary source of non-verbal communication (Cassell et al., 1994). The human-like head-and-shoulder agents do not have the same advantage.

The high comprehension scores for the female agent with the male voice were also unexpected. Together with the gender-matching voices for cartoon-agents, these yielded the highest comprehension scores. Voice quality cannot explain this effect, because there was no advantage for the male voice in male agents and the female cartoon-agent had the highest scores for the matching female voice. An explanation could lie in an unnatural voice in combination with additional social cues like gender and face, suggesting that the nature of voice is important (Mayer, Sobko & Mautone, 2003). Voice quality combined with other social cues is investigated in Experiment 2.

Experiment 2

In Experiment 2 students heard the same narration as in the previous experiment, but this time the narration was either presented by a human-agent on the screen or by no agent (voice only). The possible embodiment effect found in the previous experiment was thereby eliminated. To further investigate the interaction of the voice, but to eliminate the unnatural situation of gender-mismatching voices, we compared the effect of the voice quality of the agent using two speech synthesis systems. One system was the AT&T Labs Natural Voices Text-To-Speech System (www.research.att.com) that was used in Experiment 1. The AT&T

speech engine is a third-party product that was originally developed for and to be used with the Microsoft Speech API. It produces near-natural sounding synthetic speech. The other speech was synthesized by the Microsoft Speech Application Programming Interface (SAPI 4.0) (www.microsoft.com/speech), an engine with less superior quality but a free text-to-speech engine that works with SAPI.

Method

Participants. Thirty-two undergraduate students, 25 females and 7 males, at the University of Memphis participated in this experiment for course credit.

Materials. Materials were identical as in Experiment 1, namely the eight mini-lessons on the Roman Empire with the same multiple-choice comprehension and user impression questions following the presentations. The mini-lessons in Experiment 2 were either presented by no agent, or by a male or female agent. In the no-agent condition nothing was presented on the screen; the participant could only hear the male or female speech generated by either the AT&T or the Microsoft speech engine. In the agent condition, a human male or female agent presented the mini-lesson using either a male or female voice in one of the two speech synthesis systems.

Procedure. The procedure was identical to the one used in Experiment 1. Participants with head phone were seated in front of a computer screen and watched and listened to eight mini-lessons. After each lesson, they were asked to answer 10 comprehension and 10 impression questions. There was a 2 (presence versus absence of agent) x 2 (male /female agent and voice) x 2 (Microsoft versus AT&T speech engine) factorial design.

Results

Table 2 shows cell means and standard deviations for the comprehension scores. A repeated measures ANOVA was performed on the proportions of correct responses on the comprehension test. There was a marginally significant interaction between speech engine and gender ($F(1, 31) = 4.16, p = .058, MSE = 34.62, \eta^2 = .21$). Comprehension scores were higher for the AT&T speech engine than the Microsoft speech engine for the male-voice agents only, regardless of the presence of the agent. However, for the female agent, the opposite was true, with higher comprehension scores for the Microsoft speech engine than the AT&T speech engine when the agent was present.

 TABLE 2 ABOUT HERE

ANOVA's were performed on the ratings for impression questions. Significant differences were found for the speech engine in the perception of language use ($F(1, 31) = 16.12, p < .001, MSE = 6.88$) and life-likeness of the voice ($F(1, 31) = 19.99, p < .001, MSE = .61$), with higher ratings for the AT&T speech engine than the Microsoft speech engine ($M = 3.32, SD = .67$ vs. $M = 4.63, SD = .76$; $M = 2.95, SD = .76$ vs. $M = 2.50, SD = .81$ respectively). The impression questions revealed that participants have a preference for the higher quality AT&T speech engine. The comprehension questions showed the same pattern, barring the exception in female agents with male voices that was also found in the first experiment.

Discussion

According to the social-cue hypothesis the presence of agents with gender-matching voices should yield highest comprehension scores. Although patterns in the comprehension scores showed an advantage of the presence of the agent, this difference did not reach the significance level. Contrary to the social-cue hypothesis, no main effect in comprehension was found for the speech engine, although in the line of this hypothesis the superior speech engine was preferred overall, suggesting a discrepancy between liking and learning (Moreno, Klettke, Nibbaragandla, Graesser & TRG, 2002). This result suggests that the voice alone may be a sufficient social cue. If enough social-cues are provided by the voice only, the agent does not contribute much more to comprehension (Craig, Gholson & Driscoll, 2002; Moreno, Mayer, Spires & Lester, 2001).

The surprising finding was the interaction between speech engine and gender, with an advantage in comprehension for the presence of the female agent with the lower quality speech engine. What is so special about the female agent with the lower quality speech engine (or, for that matter, about a female agent with a male voice)? To answer this question we looked at the characteristics of the female speech engine voices by analyzing their pitch contours. A typical AT&T text-to-speech sound file with a female speaker has an F0 range from 125-250 Hz. A typical Microsoft text-to-speech sound file with a female speaker has an F0 range from 115-215 Hz. Female speakers have generally higher pitched voices as well as a wider frequency range than male speakers (Syrdal, 1996). A typical F0 range for a female speaker is in the range of 150-350 Hz, whereas a male speaker is in the range of 80-200 Hz. For female voices, the Microsoft speech engine therefore provides more male-like speech, particularly compared to the AT&T speech engine.

This would suggest that human-like female agents that sound like males rather than females yield higher scores than human-like female agents that sound like females. That result can potentially be accommodated by stereotypes. Stereotypes regarding gender and its social influence are widely reported in the social psychology literature (Eagly, 1987; Eckes & Trautner, 2000). Literature on human-computer interaction has found male voices being perceived as superior to female voices. For instance, a male synthesized voice is considered more convincing than a female voice (Nass, Moon, & Green, 1997; Reeves & Nass, 1996). The results in Experiment 1 and 2 contradicting the social-cue hypothesis may therefore support a superior male voice hypothesis based on stereotype theories in social psychology.

Conclusion

Two experiments showed that social cues like the type of agent that is used, the gender of agent and voice as well as the voice quality contribute to comprehension and liking. Moreover, there are complex interactions among these factors, a result that refutes any simple generalizations. According to available evidence, there is (a) a natural interaction with conversational agents that is on par with interactions with humans, (b) a preference for natural agents, (c) a recurrent finding that female agents with male voices yield highest comprehension scores. These findings support the general conclusion that human characteristics are applied in the perception of computational animated conversational agents.

As has been argued in the literature (Baylor & Ryu, 2003), different modalities used in these HCI studies often make findings inconclusive. For instance, the social-cue hypothesis would at least predict an effect for the presence of an animated conversational agent. We did not find conclusive evidence for such an effect. To some extent this can be explained by the experimental design we have used. At the same time, however, there is also evidence that there

is an interaction between different social cues. If the social cue of an animated agent is present, users will attend to it, but if it is absent, they will rely on the voice only (Experiment 2).

Evidence for such an interaction also comes from the stereotype effect we found: the mismatch between the gender of voice and agent depends on the non-embodied human (Experiment 1) presence (Experiment 2) of the agent.

Our findings provide further support for the social agency hypothesis and the social-cue hypothesis. Our results suggest that multiple social cues are not necessarily needed for users to interpret computers as humans. In other words, more social cues does not necessarily mean more social agency. Instead, we found that users pick up on any social cue being present to form their interpretation. If the social cue of agent is present (Experiment 1) they will pick up on it, but if it is absent (Experiment 2) there will not be large consequences (Atkinson, 2002; Craig, Gholson & Driscoll, 2002).

Despite the consistent findings in these two studies and similar findings in other studies, we need to be careful regarding any generalizations. First of all, we were not able to look at the effect of the gender of the participant because of the design of the experiment. In the first experiment gender is almost equally distributed, in the second experiment the majority of participants were female. The consistency in the findings between the two experiments does not indicate an interaction between participant gender and agent gender, but such an effect certainly can not be ruled out.

Another important point to be made is that most studies on animated conversational agents and intelligent systems rely on short-term reactions. However, animated agents in applications like intelligent tutoring systems are intended to be used for months or even years.

Whether the effects found in this study and other studies persist over time remains an open question. Practical considerations make providing an answer to this question difficult.

Based on the current experiments, what do the results say with regard to the development of animated conversational agents? Is the ideal agent a human female agent with a male-sounding voice? Findings are inconclusive: embodied cartoon-agents with gender-matching voices did equally well. Furthermore, no significant difference was found between the presence and absence of either the female agent or male agent. What we did find is that users attend to agents and apply stereotypes to animated conversational agents. It is therefore recommended to consider different social cues and their complex interactions when developing animated conversational agents. The interactions between social cues found in this study present a challenge to cross-comparisons of animated conversational agents. Future research will therefore need to identify the precise conditions when and in which social cues reign supreme.

References

- Allen, J.F., Schubert, L.K., Ferguson, G., Heeman, P., Hwang, C.H., Kato, T., Light, M., Martin, N., Miller, B., Poesio, M., & Traum, D. (1995). The TRAINS project: A case study in building a conversational planning agent. *Journal of Experimental and Theoretical AI*, 7, 7-48.
- André, E., Rist, T., & Müller, J. (1998). Integrating reactive and scripted behaviors in a life-like presentation agent. In Sycara, K. P., & Wooldridge, M. (Eds.), *Proceedings of the Second International Conference on Autonomous Agents* (pp. 261-268). Minneapolis: ACM Press.
- Atkinson, R. (2002). Optimizing learning from example using animated pedagogical agents. *Journal of Educational Psychology*, 94, 416-427.
- 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. and Thorisson, K. (1999). The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. *Applied Artificial Intelligence*, 14, 519-538.
- Cassell, J., Pelachaud, C., Badler, N.I., Steedman, M., Achorn, B., Beckett, T., Douville, B., Prevost, S. & Stone, M. (1994). Animated conversation: rule-based generation of facial display, gesture and spoken intonation for multiple conversational agents. *Computer Graphics 1994 Proceedings*, 28, 413-420.

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.

Eagly, A. (1987). *Sex differences in social behavior: A social-role interpretation*, Hillsdale, NJ: Erlbaum.

Eckes, T., & Trautner, H. M. (Eds.) (2000). *The developmental social psychology of gender*. Mahwah, NJ: Lawrence Erlbaum.

Franklin, S. (2001). Automating human information agents. In Z. Chen and L. C. Jain (Eds.), *Practical applications of intelligent agents* (pp. 27-58). Berlin: Springer.

Gertner, A., & VanLehn, K (2000). Andes: A coached problem solving environment for physics. In Gauthier, G., Frasson, C. & VanLehn, K. (Eds.), *Intelligent Tutoring Systems: 5th International Conference, ITS 2000* (pp. 131-142). Berlin: Springer.

Graesser, A. C., Hu, X., Person, N. K., Stewart, C., Toth, J., Jackson, G. T., Susarla, S., & Ventura, M. (2002). Learning about the ethical treatment of human subjects in experiments on a web facility with a conversational agent and ITS components. In S. A. Cerri, G.

Gouarderes, & F. Paraguacu (Eds.), *Intelligent tutoring systems 2002* (pp. 972-981). Berlin: Springer.

Graesser, A.C., Jackson, G.T., Ventura, M., Mueller, J., Hu, X., Person, N.K. (2003). The impact of conversational navigational guides on the learning, use, and perceptions of users of a web site. *Proceedings of the 2003 AAAI Spring Symposia on Natural Language Generation in Spoken and Written Dialogue*. (pp.9-14). Palo Alto, CA: AAAI Press.

Graesser, A.C., Lu, S., Jackson, G.T., Mitchell, H.H., Ventura, M., Olney, A., & Louwerse, M.M. (2004). AutoTutor: A Tutor with Dialogue in Natural Language. *Behavior Research Methods and Instrumentations*.

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, and J. Kay (Eds.), *Proceedings of Artificial Intelligence in Education*. (pp. 47-54). Amsterdam: IOS Press.

Graesser, A. C., Person, N., Harter, D. & The Tutoring Research Group, (2001). Teaching tactics and dialog in AutoTutor. *International Journal of Artificial Intelligence in Education*, 12, 257-279.

Graesser, A.C., VanLehn, K., Rosé, C., Jordan, P., & Harter, D. (2001). Intelligent tutoring systems with conversational dialogue. *AI Magazine*, 22, 39-51.

- Hunt, E., & Minstrell, J. (1996). Effective instruction in science and mathematics: Psychological principles and social constraints. *Issues in Education: Contributions from Education Psychology*, 2, 123-162.
- Lesgold, A., Lajoie, S. P., Bunzo, M., & Egan, G. (1992). SHERLOCK: A coached practice environment for an electronics troubleshooting job. In J. H. Larkin & R. W. Chabay (Eds.), *Computer assisted instruction and intelligent tutoring systems: Shared goals and complementary approaches* (pp. 201-238). Hillsdale, NJ: Lawrence Erlbaum.
- 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*, 23-30.
- Lester, J., Stone, B., & Stelling, G. (1999). Lifelike pedagogical agents for mixed initiative problem solving in constructive learning environments. *User Modeling User-Adapted Interaction*, 9, 1-44.
- Lewis, W. and Rickel, J.W. (2000). Animated pedagogical agents: Face to face interaction in interactive learning environments. *International Journal of Artificial Intelligence in Education*, 47-78.
- Louwerse, M. M., Graesser, A. C., Olney, A., & Tutoring Research Group (2002). Good computational manners: mixed-initiative dialogue in conversational agents. In C. Miller (Eds.),

Etiquette for human-computer work: Papers from the 2002 fall symposium, Technical Report FS-02-02 (pp. 71-76). North Falmouth: AAAI Press.

Massaro, D. W., & Cohen, M. M. (1994). Visual, orthographic, phonological, and lexical influences in reading. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 1107- 1128.

Mayer, R.E., Moreno, R., Boire, M., & Vagge, S.(1999). Maximizing constructivist learning from multimedia communications by minimizing cognitive load, *Journal of Educational Psychology*, 91, 638-643.

Mayer, R. E., Sobko, K., & Mautone, P. D. (2003). Social cues in multimedia learning: Role of speaker's voice. *Journal of Educational Psychology*, 94, 419-425.

McNamara, D. S., Levinstein, I. B. & Boonthum, C. (2004). iSTART: Interactive Strategy Trainer for Active Reading and Thinking. *Behavioral Research Methods, Instruments, and Computers*.

Moreno, K.N., Klettke, B., Nibbaragandla, K., Graesser, A.C., & the Tutoring Research Group (2002). Perceived characteristics and pedagogical efficacy of animated conversational agents. In S. A. Cerri, G. Gouarderes, & F. Paraguacu (Eds.), *Intelligent Tutoring Systems 2002* (pp. 963-971). Berlin, Springer.

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.

Nass, C., Moon, Y., & Green, N. (1997). Are computers gender-neutral? Gender stereotypic responses to computers. *Journal of Applied Social Psychology, 27*, 864-876.

Pellom, B., Ward, W. & Pradhan, S. (2000). The CU communicator: An architecture for dialogue systems. *International Conference on Spoken Language Processing (ICSLP)*. Beijing, China.

Picard, R.W. (1997). *Affective computing*. The MIT Press, Cambridge, MA.

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

Rickel, J. and Johnson, W. L. (1999). Animated agents for procedural training in virtual reality: Perception, Cognition, and Motor Control. *Applied Artificial Intelligence, 13*, 343-382.

Shneiderman, B. (1992). *Designing the user interface: Effective strategies for effective human-computer interaction*. Reading, MA: Addison-Wesley.

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.

Syrdal, A.K. (1996). Proceedings of the Fourth International Conference on Spoken Language Processing (pp. 438-441). Piscataway, NJ: IEEE.

Author Note

Max M. Louwerse, Department of Psychology / Institute for Intelligent Systems, University of Memphis; Arthur C. Graesser, Department of Psychology, University of Memphis; Shulan Lu, Department of Psychology, University of Memphis; Heather H. Mitchell, Department of Psychology, University of Memphis.

This research was supported by grants from National Science Foundation (IIS-0416128, ITR 0325428, REC0106965, SBR 9720314) and DoD Multidisciplinary University Research Initiative (MURI) program administered by the Office of Naval Research (N00014-00-1-0600) . Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of DoD, ONR, or NSF.

Correspondence concerning this article should be addressed to Max M. Louwerse, Institute for Intelligent Systems / Department of Psychology, University of Memphis, 202 Psychology Building, Memphis, Tennessee 38152-3230.

Table 1.

Proportions of Correct Mean and (SD) for Experiment 1 comprehension questions as a Function of Agent Type, Agent Gender, and Voice Gender

	Human agent		Cartoon agent		Total
	Male voice	Female voice	Male voice	Female voice	
Male agent	0.479 (.214)	0.475 (.212)	0.529 (.227)	0.473 (.203)	0.489 (.214)
Female agent	0.529 (.202)	0.483 (.184)	0.479 (.200)	0.527 (.204)	0.505 (.197)
Total	0.504 (.208)	0.479 (.198)	0.504 (.214)	0.500 (.203)	



Table 2

Proportions of Correct Mean and (SD) for Experiment 2 comprehension questions as a Function of Speech Engine, Agent and Voice Gender

	Agent absent		Agent present		Total
	Male voice	Female voice	Male voice	Female voice	
Microsoft	0.506 (.208)	0.582 (.155)	0.494 (.214)	0.629 (.176)	0.553 (.188)
AT&T	0.535 (.206)	0.565 (.183)	0.600 (.187)	0.576 (.189)	0.569 (.191)
Total	0.521 (.207)	0.574 (.169)	0.547 (.201)	0.603 (.183)	

Figure Caption Page

Figure 1. Human-like and cartoon-like male and female agents

	Female	Male
Human-like		
Cartoon-like	