Towards Caring Machines

Timothy W. Bickmore
Boston University School of Medicine
Boston, Massachusetts, USA
bickmore@bu.edu

Rosalind W. Picard
MIT Media Laboratory
Cambridge, Massachusetts, USA
picard@media.mit.edu

Abstract
The perception of feeling cared for has beneficial consequences in education, psychotherapy, and medicine. Results from a longitudinal study of simulated caring by a computer are presented, in which 60 subjects interacted with a computer agent daily for a month, half with a "caring" agent and half with an agent that did not use behaviors to demonstrate caring. The perception of caring by subjects in the "caring" condition was significantly higher after four weeks, and was also reflected in qualitative interviews with them, and in a significantly higher reported willingness to continue working with the "caring" agent. This paper presents the techniques that contributed to the increased perception of caring, and presents some of the implications of this new technology.

Categories & Subject Descriptors: H5.2 [**Information Interfaces and Presentation**]: User Interfaces—Evaluation/methodology, Graphical user interfaces, Interaction styles, Natural language, Theory and methods, Voice I/O.

General Terms: Design; Experimentation; Theory

Keywords: Caring; embodied conversational agent; affective computing; social interface; relational agent; emotion.

INTRODUCTION

Feeling cared for has profound effects on physiology, cognition and emotional state in humans. It plays an especially crucial role in the helping and medical professions and in education. According to Levinson, et al, "A growing body of literature suggests that outcomes of care are optimal when physicians address patients' emotional and personal concerns in addition to their biomedical problems. Patient satisfaction, patient adherence, and biological outcomes can be improved with a patient-centered model of care that demonstrates respect and caring for patients." [14] In education, it is known that the presence of someone who is perceived as caring can be motivating [21], and various studies have also linked caring and other qualities of interpersonal relationships between teachers and students to motivational outcomes over the long term [5].

Caring is expressed not only through speech content, but through nonverbal and paraverbal modalities including facial expression, posture, and tone and timing of speech [20]. For example, facial expressiveness alone (smiling, nodding and frowning by physical therapists) has been found to be significantly correlated with short- and long-term improvements in functioning in geriatric patients [1].

Feeling Cared For

In the same way that researchers differentiate between authentic trust and interface behaviors that elicit the perception of trust [6], we differentiate between authentic caring and interface behaviors that elicit the perception of caring. This paper focuses upon the latter, a topic for which we can draw upon numerous studies in communication, sociology and sociolinguistics to identify a set of communicative behaviors associated with caring. By implementing these behaviors in a computer interface we hypothesize that we can increase the perception of feeling cared for by users who interact with it. We also consider "feeling cared for" to be a state of subjective feeling, rather than a function of observable behavior. Moreover, the feeling is situation specific and felt with respect to another person or agent, rather than a generalized disposition.

Feeling cared for is closely associated with the concept of perceived social support, especially following earlier definitions such as "the subjective feeling of belonging, of being accepted or being loved," and is also associated with trust, when defined as "people's abstract positive expectations that they can count on partners to care for them and be responsive to their needs, now and in the future" [2].

BEHAVIORS INDICATIVE OF CARING

There are several human communicative behaviors an interface agent or robot could use to elicit the perception of feeling cared for by a user. Although providing any kind of social support can indicate caring, demonstrations of empathy and comforting behavior are perhaps the quintessential examples, and are widely cited in the helping literature as being key in achieving desired outcomes [11,20]. Other behaviors that can contribute to an

CHI 2004, April 24–29, 2004, Vienna, Austria.
ACM 1-58113-703-6/04/0004.

impression of caring include social dialogue, self-disclosure, emphasizing commonalities, meta-relational communication (particularly emotional aspects) talking about the past and future together, continuity behaviors (appropriate greetings and farewells and talk about the time spent apart), and reference to mutual knowledge, as well as explicit messages of esteem (see [3] for a summary).

As mentioned above, there are also nonverbal behaviors indicative of caring such as facial expressiveness (including displays of concern), head nodding, and tone and timing of speech. Nonverbal "immediacy" behaviors—including close conversational distance, direct body and facial orientation, forward lean, increased and direct gaze, frequent gesturing and postural openness—have been found to project liking for the other and engagement in the interaction, and to be indicative of caring [19]. Of course, to use these behaviors a computer (or robot) needs to have a body capable of conveying these human-like behaviors [7].

PREVIOUS WORK

The most relevant work to date on computer caring is the CASPER affect-management agent developed by Klein [12]. The system uses active listening techniques (displays of empathy) via text menus, and was shown to be significantly more effective at alleviating computer users' frustration compared with identical systems that only allowed users to express their feelings (vent) or ignored their feelings altogether. However, subjects were not explicitly asked if they felt cared for by the computer, so it is unknown whether the computer had any impact on this.

Although there have not been any studies to date on the ability of computer agents to affect the feeling of being cared for, there have been many studies on the ability of computer agents to affect related psychosocial constructs, such as liking of and trust in the agent. Reeves and Nass demonstrated the ability of computers to increase users' liking of them through the use of flattery, praise of other computers, matching the user in personality, or the use of "in-group" cues [18]. Morkes, Kernal and Nass demonstrated that computer agents that use humor are rated as more likable, competent and cooperative than those that do not [15]. There have also been several studies on the ability of computers to affect users' trust in them. Embodied pedagogical agents, especially those that are highly expressive, have been found to increase students' perceptions of trust: such agents are perceived as helpful, believable, and concerned [13]. Mulken, et al, found that personification of an interface by itself does not appear to be a sufficient condition for raising the trustworthiness of a computer [16]. In an experiment with REA—a life-sized, animated virtual real estate agent—Bickmore showed that the agent's use of social dialogue increased trust in it for extroverts (for introverts it had no effect) [4].

Other studies have demonstrated that (at least some) people seem to form emotional bonds with computer agents and robots. In a recent study of AIBO robotic dog owners, Friedman found that several owners indicated they felt a reciprocal emotional connection with their virtual pet [9].

Finally, the University of Pittsburgh/CMU Nursebot project involves the development of an autonomous robot for eldercare that, among other functions, is intended to provide social interaction for isolated elders. In an experiment involving this robot, Kiesler and Goetz found that subjects performed more physical exercise for a serious but caring version of the robot compared to a playful version [10].

A STUDY ON THE EFFICACY OF A CARING AGENT

Evidence that computers can instill a sense of caring comes from a recently completed study on the longitudinal effects of relationship-enhancing behaviors used by a computer agent on measures of user-computer relationship quality [3]. In this study the agent—named Laura—played the role of an exercise advisor designed to help subjects through a behavior change program designed to increase their physical activity levels. The agent appeared as an embodied conversational agent [7], whose speech and nonverbal behavior (including hand gestures, eye gaze behavior, posture shifts, head nods, proximity and facial expressions) were controlled using the BEAT text-to-embodied speech engine [8] (see Figure 1). Subjects conducted a 10 minute interaction with Laura daily on their home computers for one month, during which Laura provided feedback on their exercise behavior, helped them overcome obstacles to exercise, provided educational content related to exercise, and obtained and followed up on commitments to exercise.

A RELATIONAL version of the agent used all of the caring behaviors described above. For example, if a subject indicated they were not feeling well (and thus unable to exercise), Laura provide appropriate empathetic feedback while exhibiting a concerned facial expression (as in Figure 1). A NON-RELATIONAL version of the agent delivered identical health content but had all caring and relational behaviors removed.

The principal outcome measure used in the study was the Working Alliance Inventory, a 36-item self-report questionnaire used in psychotherapy that measures the trust and belief that the therapist and patient have in each other as team-members in achieving a desired outcome [11]. The bond subscale of this instrument assesses the emotional bond between the helper and helpee and includes questions that specifically address the helpee's feeling cared for.

Thirty-three subjects completed the month of interactions with the RELATIONAL agent and twenty-seven subjects completed interactions with the NON-RELATIONAL agent. Subjects were recruited from the MIT campus, were mostly (69%) students and were 60% female (balanced across the two conditions).

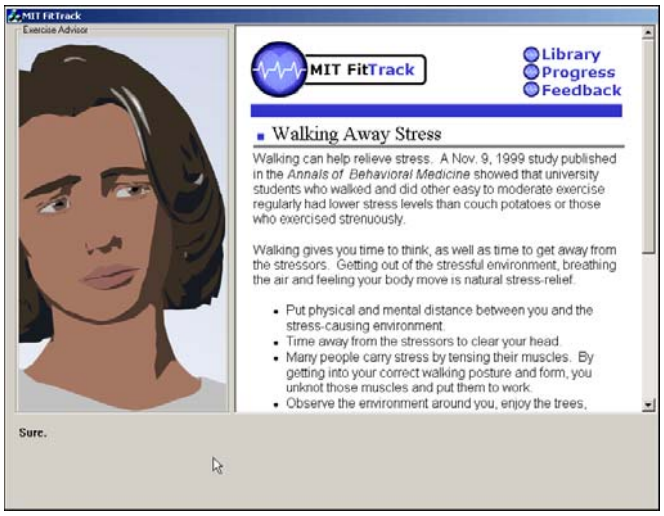

Figure 1. Exercise Advisor Agent

Quantitative Results

In this paper we only describe results that are particularly relevant to the notion of caring; for a full description see [3]. These results include the following items from the bond subscale of the Working Alliance Inventory, evaluated after four weeks of daily interaction. Subjects in the RELATIONAL condition indicated significantly greater agreement (on 7-point Likert scales) with the following items, compared with subjects in NON-RELATIONAL:

- "I feel that Laura cares about me in her own unique way, even when I do things that she does not approve of." $t(60)=2.39$, $p<.05$
- "I feel that Laura, in her own unique way, is genuinely concerned about my welfare." $t(60)=2.19$, $p<.05$
- "I feel that Laura, in her own unique way, likes me." $t(60)=2.56$, $p<.05$
- "Laura and I trust one another." $t(60)=2.05$, $p<.05$

When asked at the end of the month if they would like to continue working with Laura, subjects in the RELATIONAL condition also responded much more favorably than the NON-RELATIONAL group, $t(57)=2.43$, $p=.009$.

One behavioral measure related to caring was evaluated. In the closing session, subjects were given a choice of farewell greetings to say goodbye to the agent. Significantly more subjects in the RELATIONAL group (69%) chose the most sentimental farewell ("Take care Laura, I'll miss you." vs. "Bye.") than in the NON-RELATIONAL condition (35%), $t(54)=2.80$, $p=.004$.

Qualitative Feedback

After the experiment and before the debriefing, we asked subjects about their experiences with Laura. First, when asked whether they liked the overall concept of conversing with and relating to an animated character, subjects reported strong opinions on both sides of the issue. Representative responses included:

- "It was a really, really great idea to have some kind of animated character because it makes you feel like you're actually talking to a person rather than having words on the computer screen."
- "Personally I detested Laura."
- "I like talking to Laura, especially those little conversations about school, weather, interests, etc. She's very caring. Toward the end, I found myself looking forward to these fresh chats that pop up every now and then. They make Laura so much more like a real person."

When asked "Do you feel that she really cared about you?", many subjects responded affirmatively but qualified their responses with comments such as:

- "Yes, as much as a computer can care."
- "Yea, I think there was an illusion there that she did."
- "As much as it mattered to ... I never forgot that it was a computer program, but you'll notice that I find myself calling her by feminine pronouns rather than calling her an 'it'. So, I definitely remembered that she was a computer program, but I did feel like it was a more personal interaction than that."

Other subjects responded with uncertainty about the concept of Laura "caring":

- "I find 'care' to be a funny term to use with a computer character. I felt like it was helpful to have positive reinforcement, even if it was from a computer character."
- "She's a computer character. I don't know if she cared about me. I don't know if she feels. She's a character and has a role, but I don't know if she has feelings. But, it worked for me and I'm happy."

Finally, there was a group of subjects who answered negatively, emphasizing Laura was a machine.

- "No, not really, because I plugged in a number and she had a script."
- "No. I felt like I was talking to a robot, to a machine."

These responses illustrate a range of user feedback about a system that might evoke feelings of caring – from liking to disliking, from acceptance of the effects to denial of any effects.

CONCLUSION, IMPLICATIONS AND FUTURE WORK

This paper has described the background and motivation for building computers that elicit the perception of caring, together with a set of behaviors that were implemented and tested in a month-long study, and which resulted in a significant impact on people's perception of caring.

The results imply that the technology can influence perception of caring on a significant number of users, even when the population includes people who detest talking to agents and who are computer-savvy MIT students who know machines don't experience genuine caring. These findings are significant given that the feeling of being cared for has been widely documented to have important implications in human-human interaction, especially in education and in medicine. In addition to the benefits known to be associated with eliciting caring feelings in those domains, one can also imagine more controversial uses of this technology, perhaps to explicitly deceive somebody into thinking that they are cared for, and then to exploit them. Picard and Klein have elsewhere described several undesirable implications of technology that evinces signs of "artificial caring" [17].

The current findings are based on an agent that uses many caring behaviors in a particular task context. Follow-on studies are needed to examine subsets of these behaviors, task and other context dependencies, and comparison with other interface modalities.

ACKNOWLEDGMENTS

We thank Justine Cassell, Candace Sidner and Amanda Gruber for their contributions. This material is based upon work supported through NSF Grant 0087768.

REFERENCES

1. Ambady, N., Koo, J., Rosenthal, R., and Winograd, C. Physical Therapists' Nonverbal Communication Predicts Patients' Health Outcomes. *Psychology and Aging, 17*, 3 (2002) 443-452.
2. Berscheid, E. and Reis, H. Attraction and Close Relationships. In D. Gilbert, S. Fiske, and G. Lindzey, Eds., *The Handbook of Social Psychology*, McGraw-Hill, NY, 1998, 193-281.
3. Bickmore, T. *Relational Agents: Effecting Change through Human-Computer Relationships*. PhD Thesis, Media Arts & Sciences, MIT, Cambridge, MA, 2003.
4. Bickmore, T. and Cassell, J., Relational Agents: A Model and Implementation of Building User Trust. *Proc. CHI 2001*, CHI Letters 3(1), 396-403.
5. Birch, S. H. and Ladd, G. W., Interpersonal Relationships in the School Environment and Children's Early School Adjustment. In J. Juvonen and K. Wentzel, Eds., *Social Motivation: Understanding Children's School Adjustment*, Cambridge University Press, NY, 1996.
6. Cassell, J. and Bickmore, T. External Manifestations of Trustworthiness in the Interface. *Communications of the ACM, 43*, 12 (2000), 50-56.
7. Cassell, J., Sullivan, J., Prevost, S., and Churchill, E., Eds., *Embodied Conversational Agents*. The MIT Press, Cambridge, MA, 2000.
8. Cassell, J., Vilhjálmsson, H., and Bickmore, T., BEAT: The Behavior Expression Animation Toolkit. *Proc. SIGGRAPH 2001*, 477-486.
9. Friedman, B., Kahn, P., and Hagman, J., Hardware Companions? What Online AIBO Discussion Forums Reveal about the Human-Robotic Relationship. *Proc. CHI 2003*, CHI Letters 5(1), 273-280.
10. Goetz, J., Kiesler, S., and Powers, A., Matching robot appearance and behavior to tasks to improve human-robot cooperation., *Proc. Workshop on Robot and Human Interactive Communication: RO-MAN*, 2003.
11. Horvath, A. and Greenberg, L. Development and Validation of the Working Alliance Inventory. *Journal of Counseling Psychology, 36*, 2 (1989) 223-233.
12. Klein, J., Moon, Y., and Picard, R. This Computer Responds to User Frustration: Theory, Design, Results, and Implications. *Interacting with Computers, 14* (2002) 119-140.
13. Lester, J. C., Converse, S. A., Kahler, S. E., Barlow, S. T., Stone, B. A., and Bhogal, R. S. The Persona Effect: Affective Impact of Animated Pedagogical Agents. *Proc. CHI 1997*, 359-366.
14. Levinson, W., Gorawara-Bhat, R., and Lambs, J. A Study of Patient Clues and Physician Responses in Primary Care and Surgical Settings. *JAMA, 284*, 8 (2000) 1021-1027.
15. Morkes, J., Kernal, H., and Nass, C. Humor in Task-Oriented Computer-Mediated Communication and Human-Computer Interaction. *Proc. CHI 1998* 215-216.
16. Mulken, S. v., Andre, E., and Muller, J. An Empirical Study on the Trustworthiness of Life-Like Interface Agents. *Proc. HCI 1999*, 152-156.
17. Picard, R. and Klein, J. Computers that recognize and respond to user emotion: theoretical and practical implications. *Interacting with Computers, 14* (2002) 141-169.
18. Reeves, B. and Nass, C. *The Media Equation*. Cambridge University Press, Cambridge, 1996.
19. Richmond, V. and McCroskey, J., Immediacy. In *Nonverbal Behavior in Interpersonal Relations.* Allyn & Bacon, Boston, MA, 1995, 195-217.
20. Squier, R. A model of empathic understanding and adherence to treatment regimens in practitioner-patient relationships. *Social Science and Medicine, 30* (1990) 325-39.
21. Wentzel, K. Student Motivation in Middle School: The Role of Perceived Pedagogical Caring. *Journal of Educational Psychology, 89*, 3 (1997) 411-419.