Subtle Expressivity by Relational Agents

Timothy W. Bickmore
MIT Media Laboratory
20 Ames St. E15-120G
Cambridge, MA 02139 USA
+1 617 253 6341
bickmore@media.mit.edu

Rosalind W. Picard
MIT Media Laboratory
20 Ames St. E15-020G
Cambridge, MA 02139 USA
+1 617 253 0611
picard@media.mit.edu

ABSTRACT
Relational agents are computational artifacts designed to build long-term, social-emotional relationships with their users. In this paper we argue that subtle expressivity is especially crucial in human-computer interactions with relational agents in which social tasks such as relationship building or negotiation are being performed. We discuss these issues in the context of a relational agent designed to interact repeatedly with users during a one-month exercise adoption program.

Keywords
Subtle expressiveness, affect, emotion, non-verbal behavior, long-term interaction, embodied conversational agent

1. RELATIONAL AGENTS

Over the last three years we have begun investigating the development and use of Relational Agents; computational artifacts designed to build and maintain long-term, social-emotional relationships with their users (Bickmore, 2003). These can be purely software humanoid animated agents, but they can also be non-humanoid or embodied in various physical forms, from robots, to pets, to jewelry, clothing, hand-held, and other interactive devices. Central to the notion of relationship is that it is a persistent construct, spanning multiple interactions, thus Relational Agents are explicitly designed to remember past history and manage future expectations in their interactions with users. Finally, relationships are fundamentally social and emotional, and detailed knowledge of human social psychology--with a particular emphasis on the role of affect--must be incorporated into these agents if they are to effectively leverage the mechanisms of human social cognition in order to build relationships in the most natural manner possible.

Human relationships are primarily established in the context of face-to-face conversation. In addition to primacy of this interactional modality, face-to-face conversation affords many channels of subtle expressivity that are especially crucial in relational conversation. Thus, we have focused on developing relational agents that have anthropomorphic forms, implemented as embodied conversational agents (Cassell, Sullivan, Prevost, & Churchill, 2000)or sociable robots (Breazeal, 2002). The social and subtly expressive communicative cues used by these agents are gleaned from studies of human-human face-to-face conversation.

2. SUBTLE EXPRESSIVITY IN HUMAN RELATIONAL INTERACTIONS

Several studies have demonstrated what most business people already know: when the social aspects of an interaction are especially important--such as when you are getting to know a new client or negotiating an important deal--nothing beats face-to-face interaction. In a review of studies comparing video and audio-mediated communication, Whittaker and O'Conaill concluded that video was superior to audio only for social tasks while there was little difference in subjective ratings or task outcomes in tasks in which the social aspects were less important (Whittaker & O'Conaill, 1997). They found that for social tasks, interactions were more personalized, less argumentative and more polite when conducted via video-mediated communication, that participants believed video-mediated (and face-to-face) communication was superior, and that groups conversing using video-mediated communication tended to like each other more, compared to audio-only interactions. Obviously, some nonverbal communication must be responsible for these differences.

We define "subtle expressivity" to be those communicative behaviors used to convey any kind of meaning *except* for the primary propositional meaning of a communicative act.

The set of general functions of subtly expressive communicative behaviors studied in the literature on human-human communication is expansive, but includes:

- Interactional functions, such as: turn-taking (Duncan, 1974); engagement, disengagement, greeting and farewell (Kendon, 1990); and grounding (Clark, 1992).
- Framing (i.e., the use of "contextualization cues" to mark the type of interactional segment being initiated) (Tannen, 1993).
- Social deixis (i.e., marking relational stance) (Levinson, 1983).
- Conveying attitude (e.g., interpersonal attitude) (Argyle, 1988).

- Emphasis.
- Conveying emotional state (Argyle, 1988).

There is a correspondingly large array of communicative behaviors that have been found to be used to perform these functions, and there is a many-to-many mapping between them (a given behavior can be used to perform multiple functions and a given function can be performed by multiple behaviors). For example, emphasis can be marked using intonation, eyebrow raise, hand gesture or facial expression, whereas facial expression can be used not only for emphasis but for conveying attitude and emotional state. Further, most of these behaviors can also be used to convey non-subtle, propositional content (e.g., an isolated smile to indicate agreement).

Relatively little work has been done on studies of these behaviors in long-term interactions. It is known that entrainment (lexical, syntactic, prosodic, and postural)(Clark, 1992; LaFrance, 1982) occurs within a single interaction and likely continues to increase as a given dyad interacts over time. Gain-loss theory is a related phenomenon that posits that people who start out different but change to become more like each other over time along some trait or state dimension of personality will like each other more (Aronson & Linder, 1965). Forms of social deixis must necessarily change over time as the relationship between interlocutors evolves, and language use must take into account the increasing common ground between them as well as their shared (historical) discourse context. Relational partners also tend to develop idiomatic expressions (Bell & Healey, 1992), and it is likely that these idioms extend into the nonverbal domain of "subtle" behaviors.

3. AUTOMATICITY AND THE "SUSPENSION OF DISBELIEF"

Face-to-face conversation is hard work. Interlocutors must track task, conversational, and relational goals at varying levels of abstraction and respond to the dynamic moves of their partner by planning, re-planning and generating utterances to satisfy as many goals as possible, all within a few milliseconds (Berger, 1997; Waldron, Cegala, Sharkey, & Teboul, 1991). No wonder, then, that the production of most subtly expressive behaviors is completely automatic and unconscious (some researchers have even termed this level of interaction a "conversation between limbic systems" (Buck, 1993)).

Many researchers have argued that anthropomorphic agents must work actively to "suspend disbelief" in their users, in order for users to conduct natural, social interaction with them (Bates, 1994). We argue that this is exactly backwards. Studies by Reeves and Nass and others have demonstrated repeatedly that people respond to social cues from a computer in the same way that they respond to these cues from other people (Reeves & Nass, 1996). Further, people do this automatically and unconsciously; most people state emphatically that they would never behave according to social rules when interacting with a computer, immediately after completing an experiment in which they were observed to do just that.

Our experience has been that belief in a computer agent's acting like a person is automatic from the first moment of an interaction, and it is *this* belief which must be "suspended" by the user, when the agent fails to meet their expectations by behaving inappropriately. In a recent study of interactions with an animated real estate agent, we learned that her persona was inappropriate for the task (users rated her as unfriendly and cold) and that her nonverbal behavior was particularly inappropriate for social dialogue (users preferred conducting social dialogue with her over a telephone link, but preferred conducting real estate business "face to face") (Bickmore & Cassell, to appear). This experience taught us that, while it is easy to get users to readily engage an agent in social dialogue, it is an extremely challenging task to get the agent to maintain the illusion of human-like behavior over time; every aspect of the agent's appearance and verbal and non-verbal behavior must be correct or users will begin to discredit it.

4. LONG-TERM INTERACTIONS WITH RELATIONAL AGENTS

We have spent the last year developing and evaluating an exercise advisor system, in order to explore long-term relational interactions between people and relational agents. This system uses an embodied conversational agent who plays the role of an exercise advisor that users interact with on a daily basis during a one-month exercise adoption program. Exercise adoption was selected as a task domain because the current guidelines from the CDC and ACSM call for all Americans to accrue at least 30 minutes of moderate or better physical activity on most, if not all days of the week. This motivates a daily check in with an exercise advisor agent, thus giving the agent an opportunity to build a relationship with users over repeated interactions.

4.1 Subtle Expressivity in the Exercise Advisor

The subtle behaviors used by this agent include nonverbal markers of relational stance and framing. One of the most consistent findings on relational stance is that the use of "immediacy" behaviors--including close conversational distance, direct body and facial orientation, forward lean, increased and direct gaze, smiling, pleasant facial expressions and facial animation in general, nodding, frequent gesturing and postural openness--projects liking for the other and engagement in the interaction, and is correlated with increased solidarity (Argyle, 1988; Richmond & McCroskey, 1995). The specific relational cues implemented in the exercise advisor agent include: increased proximity, more frequent communicative headnods, eyebrow raises, and hand gestures, and less frequent gaze aways.

Based on a series of pilot studies of human fitness trainers and their clients, four conversational frames were

developed for the agent: a task frame, for information delivery; a social frame, for greetings, farewells, and social dialogue; an empathetic frame, for empathy exchanges (following Klein (Klein, Moon, & Picard, 2002)); and an encouragement frame for coaching and motivating users. Contextualization cues were primarily encoded in proximity, facial expression and prosody.

These nonverbal behaviors were implemented as extensions to BEAT, an extensible text-to-embodied-speech translator (Cassell, Vilhjálmsson, & Bickmore, 2001). The extensions were implemented in a "Stance Manager" module that takes relational stance and conversational frame as inputs, and outputs modifications to be applied to the agent's default nonverbal behavior. Figure 1 shows examples of the exercise advisor agent in various relational stances and conversational frames.

4.2 Long-Term Changes in Behavior

The exercise advisor agent changes its behavior over time as a function of the number of interactions with a subject and increasing common ground. The daily interactions with the agent are scripted using Augmented Transition Networks (Woods, 1986) and are designed to increase relational closeness over time, for example by increasing the amount and intimacy of social dialogue used. In addition, the agent learns facts about the subject (stored in a database between interactions) and modifies its future dialogues accordingly. While neither of these long-term adaptations directly impact the subtle behaviors described in the previous section, they do change the frequency with which different conversational frames are used. Further, social dialogue itself may be viewed as a type of subtle behavior ("phatic communion" being the best exemplar (Malinowski, 1923)) in that little propositional meaning is typically conveyed in this frame.

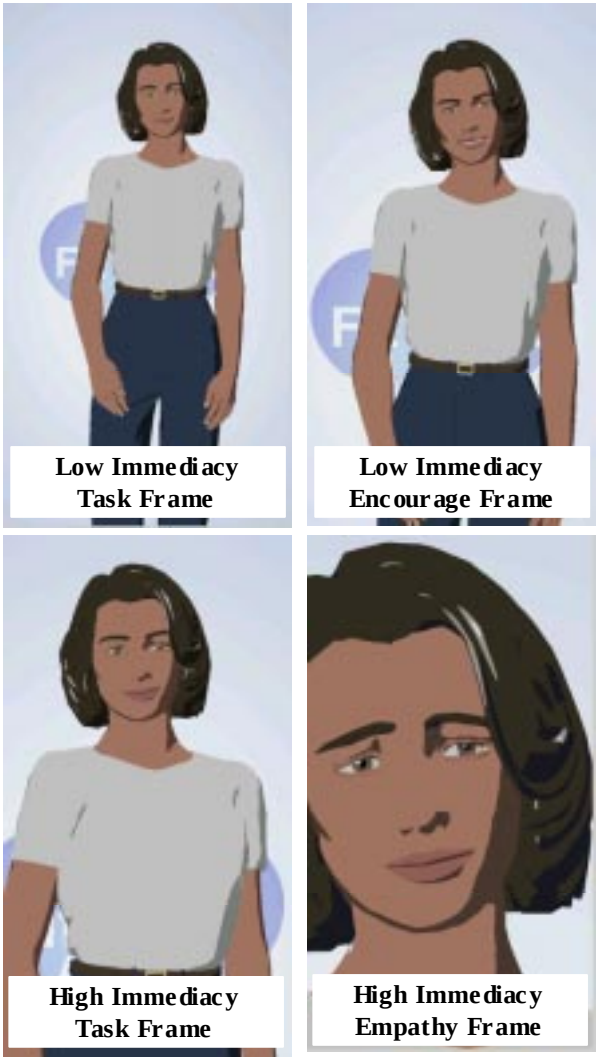

Figure 1. Effects of Relational Stance and Frame on Proximity and Facial Expression

5. EVALUATION

Evaluation of subtle expressivity can take place on multiple levels. First, evaluations can be performed that determine whether users can correctly perceive the expressive behaviors or not. Along these lines, a series of surveys was conducted on the Exercise Advisor agent to determine which nonverbal behaviors (postures and facial expressions) conveyed the intended relational stance and emotional displays.

Second, the impact of subtly expressive behaviors on user's attitudes towards the agent and the interaction can be assessed using self-report instruments and behavioral measures. A large-scale evaluation of the exercise advisor agent was recently completed, in which 100 users interacted with it on a daily basis for a month. The study used a between-subjects experimental design, with differences between two of the conditions intended to demonstrate the efficacy of long-term relationship-building strategies used by the agent. Results indicate that when the agent used these relational strategies, users reported liking, trusting and respecting the agent more, feeling that it liked, trusted, respected and cared about them more, and an increased desire to continue working with it, compared with users in the non-relational condition. The primary instrument used to assess these relational effects was the Working Alliance Inventory (Horvath & Greenberg, 1989), and we found that relationship building strategies resulted in significantly greater ratings on the bond dimension of this scale on day 7 ($t(69)=2.10$, $p<.05$) and on day 27 ($t(60)=2.54$, $p=.007$) of the intervention.

Finally, and most importantly, the impact of subtly expressive behaviors on task outcome should be measured. In the Exercise Advisor study we measured levels of physical activity through both self-report questionnaire and pedometer readings, for results refer to (Bickmore, 2003).

6. CONCLUSION

This paper has argued that Relational Agents, which build long-term social-emotional relationships with their users, need to *appropriately* employ subtle expressive capabilities. A relational agent with several subtle expressive capabilities has been designed, built and tested with over a hundred users, and shown to increase their liking of, trusting in and respecting of the agent, their feeling that it liked, trusted, respected and cared about them more, and an increased desire to continue working with it, relative to users who interacted with a non-relational agent.

ACKNOWLEDGMENTS

Thanks to Justine Cassell, Amanda Gruber, Candy Sidner, and the many folks who contributed to the development and evaluation of the Exercise Advisor system.

References

1. Argyle, M. (1988). Bodily Communication. New York: Methuen & Co. Ltd.
2. Aronson, E., & Linder, D. (1965). Gain and loss of esteem as determinants of interpersonal attractiveness. Journal of Experimental and Social Psychology, 1, 156-171.
3. Bates, J. (1994). The role of emotion in believable agents. Communications of the ACM, 37(7), 122-125.
4. Bell, R., & Healey, J. (1992). Idiomatic communication and interpersonal solidarity in friends' relational cultures. Human-Communication-Research, 18(3), 307-335.
5. Berger, C. (1997). Planning Strategic Interaction. Mahwah, NJ: Lawrence Erlbaum Associates.
6. Bickmore, T. (2003). Relational Agents: Effecting Change through Human-Computer Relationships. MIT, Cambridge, MA.
7. Bickmore, T., & Cassell, J. (to appear). Social Dialogue with Embodied Conversational Agents. In N. Bernsen (Ed.), Natural, Intelligent and Effective Interaction with Multimodal Dialogue Systems. New York: Kluwer Academic.
8. Breazeal, C. (2002). Designing Sociable Robots. Cambridge, MA: MIT Press.
9. Buck, R. (1993). The spontaneous communication of interpersonal expectations. In P. D. Blanck (Ed.), Interpersonal expectations: Theory, research, and applications (pp. 227-241). New York: Cambridge University Press.
10. Cassell, J., Sullivan, J., Prevost, S., & Churchill, E. (2000). Embodied Conversational Agents. Cambridge: MIT Press.
11. Cassell, J., Vilhjálmsson, H., & Bickmore, T. (2001). BEAT: The Behavior Expression Animation Toolkit. Paper presented at the SIGGRAPH '01, Los Angeles, CA.
12. Clark, H. H. (1992). Arenas of Language Use. Chicago, IL: University of Chicago Press.
13. Duncan, S. (1974). On the structure of speaker-auditor interaction during speaking turns. Language in Society, 3, 161-180.
14. Horvath, A., & Greenberg, L. (1989). Development and Validation of the Working Alliance Inventory. Journal of Counseling Psychology, 36(2), 223-233.
15. Kendon, A. (1990). A Description of Some Human Greetings, Conducting interaction: Patterns of behavior in focused encounters (pp. 153-207). Cambridge: Cambridge University Press.
16. Klein, J., Moon, Y., & Picard, R. (2002). This Computer Responds to User Frustration: Theory, Design, Results, and Implications. Interacting with Computers, 14, 119-140.
17. LaFrance, M. (1982). Posture Mirroring and Rapport. In M. Davis (Ed.), Interaction Rhythms: Periodicity in Communicative Behavior (pp. 279-298). New York: Human Sciences Press, Inc.
18. Levinson, S. C. (1983). Pragmatics. Cambridge: Cambridge University Press.
19. Malinowski, B. (1923). The problem of meaning in primitive languages. In C. K. Ogden & I. A. Richards (Eds.), The Meaning of Meaning: Routledge & Kegan Paul.
20. Reeves, B., & Nass, C. (1996). The Media Equation. Cambridge: Cambridge University Press.
21. Richmond, V., & McCroskey, J. (1995). Immediacy, Nonverbal Behavior in Interpersonal Relations (pp. 195-217). Boston: Allyn & Bacon.
22. Tannen, D. (1993). Introduction (Framing in Discourse). In D. Tannen (Ed.), Framing in Discourse (pp. 3-13). New York: Oxford University Press.
23. Waldron, V. R., Cegala, D. J., Sharkey, W. F., & Teboul, B. (1991). Cognitive and tactical dimensions of conversational goal management. In K. Tracy & N. Coupland (Eds.), Multiple goals in discourse (pp. 101-119). Clevedon: Multilingual Matters.
24. Whittaker, S., & O'Conaill, B. (1997). The Role of Vision in Face-to-Face and Mediated Communication. In K. Finn & A. Sellen & S. Wilbur (Eds.), Video-Mediated Communication (pp. 23-49): Lawrence Erlbaum Associates, Inc.
25. Woods, W. A. (1986). Transition Network Grammars for Natural Language Analysis. In B. J. Grosz & K. S. Jones & B. L. Webber (Eds.), Readings in Natural Language Processing (pp. 71-88). Los Altos, CA: Morgan Kaufmann Publishers, Inc.