External Manifestations of Trustworthiness in the Interface

Justine Cassell, Timothy Bickmore

Gesture and Narrative Language Group
MIT Media Laboratory
E15-315
20 Ames St., Cambridge MA 02139
{justine, bickmore}@media.mit.edu

Introduction

This article is about the kind of trust that is demonstrated in human face-to-face interaction, and approaches to and benefits of having our computer interfaces depend on these same manifestations of trustworthiness. In making technology that is actually trustworthy your morals can really be your only guide. But, assuming that you're a good person, and have built a technology that does what it promises, or that represents people who do what they promise, then read on. We're taking as a point of departure our earlier work on the effects of representing the computer as a human body. Here we are going to argue that interaction rituals among humans, such as greetings, small talk and conventional leavetakings, along with their manifestations in speech and in embodied conversational behaviors, can lead the users of technology to judge the technology as more reliable, competent and knowledgeable – to trust the technology more.

Trust is essential for all kinds of interpersonal interactions; it is the loom on which is woven the social fabric of society. Trust between humans has to do with credibility, with believing one another, with confidence in another's judgments, and beliefs that another's actions fit our own schemata of how to act. We use the interaction rituals of conversation, in part, to demonstrate our trustworthiness and, more generally, to establish and maintain social relationships where trust is important. Building rapport and common ground through small talk, intimacy through self-disclosure, credibility through the use of technical jargon, social networks through gossip, and "face" through politeness, are all examples of this phenomenon. These social uses of language are not important just in purely social settings, but are also crucial to the establishment and maintenance of any collaborative relationship or task accomplishment.

Not just does conversation demonstrate trustworthiness, but trustworthiness has an effect on how we act and converse with one another. When we do not trust, we do not believe what others say to us, nor learn from them, nor do we engage in financial or emotional transactions nor allow ourselves to disclose personal information and become more intimate. In fact, even at the local level, interactions between two people who do not trust one another are difficult to sustain: they display less verbal fluency, are filled with pregnant pauses, with incoherent sounds, with dropped words (Prentice, 1975) as each participant estimates what it's safe to reveal.

Many of the metaphors we use in American culture to signify trust derive from these kinds of interaction ritual conversational behaviors, particularly concerning the role of the human body in face-to-face interaction: "he looked me right in the eye," "he looked at her with trusting eyes", "we stood toe-to-toe", "we worked shoulder-to-shoulder", " it was service with a smile" and, conversely, "he went behind my back".

These metaphors reflect our instinctive belief that it is easiest to gauge the trustworthiness of another when we can engage in interaction rituals firsthand. What else would explain the fact that initial business meetings are still routinely held in person when the information content could easily be handled via a teleconference or videoconference? Looking people in the eye, shaking their hands, and watching them make presentations seem to be pre-requisites to establishing a working level of trust in the business world.

Caveats

In this article we are not addressing the issue of why technologies should be trustworthy, and how to make them so. We are discussing, instead, how to inspire a cognitive state of trust in the user of a technology by the use of trust-inspiring procedures: the *external manifestations of trust*, or signals of trustworthiness. In fact, as each technology that mediates communication across space and time (the telephone, e-mail, the fax) has developed, users of the technology have addressed a similar issue. Conversations that are mediated by various technologies have to overcome the lack of face-to-face data that allows participants to gauge credibility, and conversants try to find ways to overcome the lack, or to exploit it. In the days of face-to-face and door-to-door sales, Fuller Brush salesmen knew that eye contact and a winning smile would guarantee the door not closing in their face. Telemarketers soon learned to address victims by their names, and engage in small talk ("and how are you today Mrs. Brown?). E-mail marketing today attempts to personalize notices with knowledge of the person's past behavior ("since last month when you bought a copper pot with us, lots of new and exciting goods have come in of interest to you"). The current research on the potential of interaction rituals to increase trust will no doubt also have its misusers. Our belief, however, is that interaction rituals such as these *acknowledge* the social attributions that users make, and build on them to make technology that is easy to understand, quick to acquire knowledge of, and congruent in how it advertises its capabilities. We do not believe that these signals of trustworthiness can be used to trick users out of their credit card numbers with any more success than a stranger on the street might have. Rather they signal that the technology is a competent and cooperative interactant with respect to the user's sociolinguistic background; a basis for *evaluating* the level of trust one wishes to entertain with a stranger on the street.

In what follows we discuss how the interaction rituals of conversation, both verbal and nonverbal, can promote a look of trustworthiness and, in so doing, can convince users to engage in interaction in the first place and then allow them to interact in a relaxed and therefore successful manner. We want users to believe the information that the technology provides, and to reciprocate by providing information that the technology requires. How do we allay mistrust in order to allow the process of an interaction to be successful? In the domain of interpersonal trust, a useful distinction can be made between a cognitive state of trust, and trusting behaviors (Pearce, 1974). The former involves a state of (a) perceiving another person as knowing the nature of the type of interaction about to transpire, (b) able to perform his/her end of things, (c) and responsible enough to try to ensure that the interaction does not result in negative consequences for the trusting person. Trusting behaviors involve making oneself vulnerable to another person in any one of a number of ways. Our goal is to inspire a cognitive state of trust in users such that they will engage in trusting behaviors that allow the human – computer interaction to proceed smoothly.

Embodied Conversational Agents

In the Gesture and Narrative Language Group at the MIT Media Lab, we have been developing Embodied Conversational Agents (ECAs) for the last several years. ECAs are graphical computer characters that are able to engage in face-to-face dialog with a user, using not only speech but also nonverbal modalities such as gesture, gaze, intonation and posture. We have constructed several such multimodal systems that sense the user's speech, gesture, body posture and intonation and respond by animating a computer character with behavior based on studies of human conversation.

In a previous issue of CACM we wrote about the important *affordances* of the body for signaling conversational process – for indicating where in an interaction one is, whether the computer has understood the user's input, and the important uses to which embodied interfaces can be put for facilitating certain kinds of human – computer interaction (Cassell 2000). Recently, we have begun modeling some of the more social cues which people use to signal trust in face-to-face encounters, and have begun experiments into the ability of these interfaces to engage users' trust. We believe that embodied interfaces that display these social cues, and engage in the interactional rituals that display them, may elicit users' trust, allowing the interaction to manifest all of the smoothness, lack of hesitancy, and increased self-disclosure that accompanies trustful encounters among humans.

One conceptualization of the cognitive state of trust is that it is a composite of benevolence (belief in the intentional good will of another) and credibility (dis-belief in the unintentional ill will of another) (Doney and Cannon 1997). ECAs can demonstrate benevolence by engaging in interaction rituals such as greetings and small talk, relating past experiences of benevolent behavior, or referring to third-party affiliations. Interaction rituals such as these also fit into the uncertainty reduction model of trust, in which individuals incrementally reinforce their assumptions about their partner's dependability with actual evidence from their partner's behavior (Berscheid and Reis 1998). The natural progression of a conversation between strangers from greetings, through small talk, into more substantive topics can be seen as a process in which they iteratively "test the water" to determine if they want to continue deepening the relationship or not. Thus, an ECA can provide a natural transition into a trust relationship, especially for web sites or software products that users have not seen before.

ECAs can also project credibility via verbal and nonverbal modalities by presenting themselves as competent, fluid speakers, and through appearance--projecting expertise, professional affiliation, or attractiveness.

Interaction Rituals by ECAs

We are currently investigating the use of interaction rituals to build trust in ECAs carrying out real estate sales encounters. Within such encounters, with an ECA named Rea (Real Estate Agent) playing the role of a real estate agent, we believe that such interaction rituals can help the agent achieve its goals by "greasing the wheels" of task talk. It can serve a transitional function, providing a ritualized way for people to move into conversation in what may be an otherwise awkward or confusing situation (Schneider, 1988). Small talk in particular can also serve an exploratory function by providing a conventional mechanism for users to establish the capabilities and credentials of the agent (and vice-versa). Small talk can build solidarity with

users if agents engage in a ritual of showing agreement with and appreciation of user's utterances (Schneider 1988). Finally, an ECA can use a kind of small talk called conversational storytelling to establish its expertise, by relating stories of past successful problem-solving behavior, and to obtain information about the user that can be used indirectly to help achieve task goals (e.g., finding out that the user drives a minivan increases the probability that he has children).

In particular, Rea plans and carries out small talk

- as a way of heightening solidarity before face-threatening (sensitive) task topics are undertaken. Rea continually assesses her solidarity with the user, modeled as a scalar quantity. Each conversational topic has a pre-defined, pre-requisite solidarity that must be achieved before Rea can introduce the topic. Given this, the system can plan to perform small talk in order to "grease the tracks" for task talk, especially about sensitive topics like the user's income.
- as a way of transitioning from one phase of the interaction to the next, and moving towards topics that are important. Rea keeps track of the current and past conversational topics. Conversational bids which stay within topic (maintain topic coherence) are given preference over those that do not. In addition, Rea can plan to execute a sequence of conversational turns which gradually transition the topic from its current state to one that Rea wants to talk about (e.g., from talk about the weather, to talk about Boston weather, to talk about Boston real estate).
- to establish her expertise and her limits. Rea begins with some self-disclosing statements that indicate some limiting features of her technology. During the interaction she mentions other clients who resemble the user in some way, and states that she has helped them accomplish their goals.
- to acquire information about the user. Rea has a list of prioritized goals to find out about the user's housing needs in the initial interview. Conversational turns which directly work towards satisfying these goals (such as asking interview questions) are preferred.

Below is an actual instance of Rea engaging in small talk at the beginning of a real estate interview:

.   *Rea*: That microphone is terrible, I hate using those things.
.   *Rea*: Sorry about my voice, this is some engineer's idea of natural sounding.
.   *Rea*: Are you one of our sponsors?
.   *User*: Yes.
.   *Rea*: Cool.
.   *Rea*: Were you at our last sponsor meetings?
.   *User*: Yes I was. I only stayed for the first two days though.
.   *Rea*: I see.
.   *Rea*: I got so exhausted at the last sponsor meeting I think I was starting to lose my voice by the end.
.   *User*: I bet!
.   *Rea*: So, where would you like to live?
.   *User*: Cambridge
.   *Rea*: How many bedrooms do you need?
.   *User*: Two.
.   *Rea*: Do you need access to the subway?
.   *User*: Yes, please.
.   *Rea*: Is one bath enough?
.   *User*: That's fine.
.   *Rea*: You know, Boston is certainly more expensive than it used to be.
.   *User*: Oh, the Bay Area has become much more expensive too.
.   *Rea*: Really?
.   *Rea*: Anyway, what can you afford?
.   *User*: Two hundred thousand.
.   *Rea*: What kind of down payment can you make?
.   *User*: Fifteen percent.
.   *Rea*: Let me see what I have available.

We are currently conducting an experiment to assess the efficacy of this kind of small talk in real estate interviews with an ECA, and its effect on the user's perception of and trust in the agent. Subjects are interviewed by the agent about their housing needs, shown two "virtual" apartments, and then asked to fill out a questionnaire, which includes questions about how much rent they would be willing to pay for one of the apartments, attitudinal measures of perceived likability, competence, and intelligence of the agent, and a standard measure of trust. Preliminary results indicate that users find the agent more competent, reliable and knowledgeable when it uses small talk than when it engages in task talk only.

Story Listening ECAs

One important aspect of trust that we described above is its role in encouraging self-disclosure, or talking about the self. Along these lines, we are investigating the use of ECAs to prompt for and listen to the personal narratives of the user. In the Sam project (Cassell, 1999), an ECA encourages young children to engage in storytelling. In the GrandChair project, an ECA who appears to be a young child in a rocking chair listens to grandparents' family stories (Smith, 2000). The stories are videotaped so that they can be watched by future generations. In a preliminary experiment we have found that grandparents talk significantly longer in the presence

of an ECA than when using a video camera and cue cards alone. We hypothesize that this difference is at least partly due to the ECA's self-disclosure prompts (shown by other researchers to induce self-disclosure to a computer (Moon 1999), and partly due to the benevolent image and voice of a child which increases the user's trust in the system. Trust in these kinds of systems is especially crucial, since users are being asked to disclose very intimate information about themselves to a computer.

Mediating Human Interpersonal Interactions using ECAs

So far we've described instances where the embodied interface is representing the knowledge and expertise of a computational system. In those instances in making the interface look trustworthy, we're representing the trustworthiness of the computational system. But what about cases where the technology is mediating an interaction between two users, as in videoconferencing or chat. Here the issue of trust arises as users attempt to ascertain if they can trust the representation of the other human. In this domain we have developed semi-autonomous ECA avatars to represent users in interactions with other users in graphical chat systems. In these systems users control the content of what their avatar says and some aspects of their avatar's movements (walking, for example), while much of the nonverbal conversational behavior displayed by the avatar is automatically generated based on the conversational context. For example, if a user indicates that she wants to talk to another user in the chat system, her avatar will automatically produce the appropriate eye gaze, facial expression, and gestural behaviors required to signal that it wants to engage the other user in a conversation.

This kind of system gives users a higher bandwidth (more modalities) to use social cues to signal their intent to trust and be trusted, while still allowing them to maintain anonymity if they desire. In addition, semi-autonomous ECAs can actually *decrease* deception (untrustworthy behavior), by ensuring that all communicative modalities are delivering a consistent message, an important consideration in designing a system to be trusted (Nass, Isbister et al. 1999). In contrast, current graphical avatars can be controlled so that their verbal and nonverbal behavior is completely independent (e.g., someone's avatar can be smiling while flaming at you). Evaluation of this system showed that users felt better understood, felt they better understood other users, and felt

that both they and their conversational partners were more expressive when their avatars autonomously generated interaction ritual behavior such as greetings, turntaking and leavetakings, than when they directly manipulated the behaviors of their avatars (Cassell and Vilhjálmsson 1999).

Conclusion

Some degree of trust is required to engage in cooperative behavior. Conversation, in particular, requires cooperation and mutual trust to function smoothly--trust that one's partner is being truthful, is not withholding important information or conveying only superfluous or redundant information, and trust that one's partner will not blatantly insult or infringe upon one's freedom (Goffman 1959). In turn, trust may be effectively established using the same myriad social cues that people use in face-to-face conversation--using interaction rituals such as small talk to incrementally build evidence of the conversational agent's good will and credibility.

We have found that building ECAs that can engage in phatic, or relationship-oriented, behaviors challenges our notions of technology as tool, pushes us further than we expected into the metaphor of computer as conversational partner, and even provides technical challenges as we attempt to model the kinds of *goals* that small talk will achieve, in order to plan the interaction with a user. In the beginning we were even a little hesitant about admitting that our ECAs were engaging in small talk with users. Weren't we pulling some kind of *trick* on users by having our system act so much like a human, down to the very interaction cues that signal a relationship bond? In fact, as we watched children and adults interact with the system we became convinced that nobody was fooled – nobody was going to leave thinking that this was a new living species, or a new kind of human. On the other hand, in a myriad of subtle ways, users felt *heard*, felt as if the technology was adapting to them, rather than the other way around. And in continuing the project we've discovered that, unlike uses of cute small talk-like behaviors in other systems, our approach to small talk is actually responsive to user state and to the overarching goals of the interaction, and is a prime way for the system to monitor the user's progress in his or her chosen task, and to induce the user to speak easily without embarrassed pauses or disfluencies. All good reasons for users to be more trusting.

References

Berscheid, E. and H. Reis (1998). Attraction and Close Relationships. The Handbook of Social Psychology. D. Gilbert, S. Fiske and G. Lindzey. New York, McGraw-Hill.

Cassell, J. (2000). "More Than Just Another Pretty Face: Embodied Conversational Interface Agents." CACM. April, 2000.

Cassell, J., Ananny, M., Basu, A., Bickmore, T., Chong, P., Mellis, D., Ryokai, K., Vilhjálmsson, H., Smith, J., Yan, H. (2000) "Shared Reality: Physical Collaboration with a Virtual Peer." Proceedings of CHI. April 4-9, Amsterdam, NL.

Cassell, J. and H. Vilhjálmsson (1999). "Fully Embodied Conversational Avatars: Making Communicative Behaviors Autonomous." Autonomous Agenst and Multi-Agent Systems **2**: 45-64.

Doney, P. and J. Cannon (1997). "An Examination of the Nature of Trust in Buyer-Seller Relationships." Journal of Marketing **61**: 35-51.

Goffman, E. (1959). The Presentation of Self in Everyday Life New York, Doubleday.

Moon, Y. (1999). When the Computer is the "Salesperson": Consumer Responses to Computer "Personalities" in Interactive Marketing Situations. Harvard Business School 99-041.

Nass, C., K. Isbister, et al. (1999). Truth is Beauty: Researching Embodied Conversational Agents. Embodied Conversational Agents. J. Cassell et al. (eds.). Cambridge: MIT Press.

Pearce, W.B. (1974) Trust in interpersonal communication. Speech Monographs, 41(3): 236-244.

Prentice, D.S. (1975) The Effects of Trust-Destroying Communication on Verbal Fluency in the Small Group. Speech Monographs, 42(4): 262-270.

Schneider, K. P. (1988). Small Talk: Analysing Phatic Discourse Marburg, Hitzeroth.

Smith, J. (2000) GrandChair: Conversational Collection of Grandparents' Stories Using an Autonomous Animated Character. MIT Media Lab MA Thesis.