

Relational Agents: A Model and Implementation of Building User Trust

Timothy Bickmore, Justine Cassell

MIT Media Lab

20 Ames St., E15-320

Cambridge, MA 02139 USA

+1 617 253 7368

{bickmore, justine}@media.mit.edu

ABSTRACT

Building trust with users is crucial in a wide range of applications, such as advice-giving or financial transactions, and some minimal degree of trust is required in all applications to even initiate and maintain an interaction with a user. Humans use a variety of relational conversational strategies, including small talk, to establish trusting relationships with each other. We argue that such strategies can also be used by interface agents, and that embodied conversational agents are ideally suited for this task given the myriad sociocultural cues available to them for signaling trustworthiness. We describe a formal theory of social dialogue, a real implementation in an embodied conversation agent, and an experiment in which the use of social dialogue was demonstrated to have an effect on trust for users with a disposition to be extroverts.

Keywords

Embodied conversational agent, trust, social interface, natural language, small talk, personality.

INTRODUCTION

Humans are able to use a variety of strategies to proactively establish and maintain social relationships with each other. Building rapport and common ground through small talk, intimacy through self-disclosure, credibility through the use of expert's jargon, social networks through gossip, and "face" through politeness are all examples of this phenomenon. These relational strategies are important not just in purely social settings, but are also crucial to the establishment and maintenance of any collaborative relationship.

Computer interface agents may also profitably use relational strategies such as these if they are to function successfully in roles which require users to interact with them for more than a few minutes, or in which we expect users to take them seriously enough to discuss their medical problems or give out their credit card numbers. Agents of this sort must be able to establish social relationships with users in order to engage their trust which, in turn, eases cooperation.

Existing "social" interface agents (e.g., Microsoft "Bob" or the Paper Clip) achieve their social effects by attempting to draw the user into what is billed as a social interaction;

essentially a passive strategy for relationship building. What these systems lack are explicit behaviors, protocols and strategies for building, maintaining or changing a relationship with the user, something humans have a large repertoire of techniques for. Further, these systems make poor use of the primary modality humans use to establish and maintain relationships, namely language.

Embodied Conversational Agents (ECAs) are particularly well suited to the task of relationship building. ECAs are anthropomorphic interface agents which are able to engage a user in real-time, multimodal dialogue, using speech, gesture, gaze, posture, intonation, and other verbal and nonverbal channels to emulate the experience of human face-to-face interaction [10]. The nonverbal channels are important not only for conveying information (redundantly or complementarily with respect to the speech channel), but for regulating the flow of the conversation. These nonverbal channels are also especially crucial for relational conversation, since they can be used to provide such social cues as attentiveness, positive affect, and liking and attraction, and to mark shifts into and out of relational activities.

In this paper we will discuss a model of social dialogue for building user trust: we will talk about the conversational strategies that comprise the model, and one kind of talk -- small talk—that executes those strategies. Finally, we will describe an evaluation of our approach where users interacted with one of two embodied conversational agents, and we later evaluated their trust in the interaction. We concentrate on the relational notion of trust because it is essential for all kinds of interpersonal interactions, and crucially important for certain types of human-computer interactions [14]. Trust between humans involves credibility, believing one another, confidence in another's judgments, and belief that another's actions fit our own schemata of how to act [6]. Trust is a prerequisite for actions involving another agent in which one may suffer physical, financial or psychological harm (e.g., financial transactions, or disclosing personal information [31]).

Related Work in Relational Agents

In a series of studies in the "Computers As Social Actors" paradigm, researchers have demonstrated the possibility of manipulating the user's relationship with a computer using a

wide range of behaviors. Reeves & Nass demonstrated that users like computers more when the computer flatters them [23]. Morkes, Kernal and Nass demonstrated that computer agents which use humor are rated as more likable, competent and cooperative than those that do not [20]. Moon demonstrated that a computer which uses a strategy of reciprocal, deepening self-disclosure in its (text-based) conversation with the user will cause the user to rate it as more attractive, divulge more intimate information, and become more likely to buy a product from the computer [21].

Of course the social influence strategies of relational agents may not be equally effective across all types of users. Several studies have shown that users react differentially to social agents based on their own personality and other dispositional traits. For example, Reeves and Nass have shown that users like agents that match their own personality (on the introversion/extraversion dimension) more than those which do not, regardless of whether the personality is portrayed through text or speech [23] [22]. Resnick and Lammers showed that in order to change user behavior via corrective error messages, the messages should have different degrees of "humanness" depending on whether the user has high or low self-esteem ("computer-ese" messages should be used with low self-esteem users, while "human-like" messages should be used with high-esteem users) [24]. Rickenberg and Reeves showed that different types of animated agents affected the anxiety level of users differentially as a function of whether users tended towards internal or external locus of control [26].

Embodied Conversational Agents

Work on the development of ECAs, as a distinct field of development, is best summarized in [10]. In addition to REA [7] (described below), some of the other major ECA systems developed to date are Steve [25], the DFKI Persona [1], Olga [3], Gandalf [30], and pedagogical agents developed by Lester, et al, [17, 18]. There are also a growing number of commercial ECAs, such as those developed by Extempo, Headpedal, and Artificial Life, and the Ananova newscaster developed by Ananova, Ltd. These systems vary greatly in their linguistic capabilities, input modalities (most are mouse/text/speech input only), and task domains, but all share the common feature that they attempt to engage the user in natural, full-bodied (in some sense) conversation. Although these systems hold out the promise of increased engagement and effectiveness, evaluations of their use in domains from learning and training to entertainment and communication have not proved their worth. Dehn and van Mulken [13], specifically examining evaluations of recent animated interface agents, conclude that the benefits of these systems are arguable in terms of user performance, engagement with the system, or even attributions of intelligence. However, they go on to point out that virtually none of the systems evaluated exploited the human bodies they inhabited: this design paradigm "can only be expected to improve human-computer interaction if it shows some behavior that is

functional with regard to the system's aim." In light of these results, we have designed an embodied conversational agent that is based on a model of social dialogue for building user trust and diminishing interpersonal distance, and that is implemented in a domain in which exactly these abilities are key.

A MODEL OF SOCIAL DIALOGUE FOR BUILDING USER TRUST

One of the prominent theories in social psychology which claims to account for the establishment and growth of interpersonal relationships is social penetration theory [2]. In this model, relationships develop through the reciprocal exchange of information, beginning with relatively non-intimate topics and gradually progressing to more personal and private topics. Thus this theory provides an account of relationships that primarily focuses on information exchange, and the growth of a relationship can be represented in both the breadth and depth (public to private) of information disclosed. Some researchers have labeled this dimension of a relational model *familiarity* [29].

Two other dimensions of a relational model--*power* and *solidarity*--have been dealt with both in social psychology and within linguistics to account for the usage of different forms of pronouns of address (T-forms vs. V-forms [5]). Power is the ability for one interactant to control the behavior of the other. Solidarity is defined as "like-mindedness" or having similar behavior dispositions (e.g., similar political membership, family, religions, profession, sex, etc.), and is very similar to Brown and Levinson's notion of social distance in their theory of politeness [4]. There is a correlation between frequency of contact and solidarity, but not necessarily a causal relation [4, 5].

One model of the development of trust among people describes it as "a process of uncertainty reduction, the ultimate goal of which is to reinforce assumptions about a partner's dependability with actual evidence from the partner's behavior" [2]. Disclosing information to another is a behavior that communicates that we trust that person to respond appropriately. Thus, trust is predicated on solidarity and familiarity, but also likely includes self-disclosure as well as other specific trusting behaviors.

Conversational Strategies for Changing Interpersonal Relationships

Our objective is to build an ECA that knows how to win people's trust and that goes about the process using relational conversational strategies. This requires a model of trust that is broken down into the goals to be achieved and the conversational strategies for achieving them, as well as the ways of generating those conversational strategies and putting them into practice. In this section we explain two broad categories of conversational strategy that play a role in achieving increased trust -- facework, and establishing common ground. We will then turn to how these strategies can be generated and put into practice in small talk generated by an ECA.

In Goffman's dramaturgical approach to social interaction, he defined an interactant's "line" as the patterns of action by which individuals in an interaction present an image of themselves and the situation, that is their social role in the current joint activity [15]. The notion of "face", Goffman went on to say, is "the positive social value a person effectively claims for himself by the line others assume he has taken during a particular contact". Interactants maintain face by having their line accepted and acknowledged by their interactants. Events which are incompatible with their line are "face threats" and are mitigated by various corrective measures if they are not to lose face. In short, events which are incompatible with how we wish others to see us, are called "face threats", and we try to avoid them, and to mitigate their effect if they are unavoidable.

Brown and Levinson extended Goffman's notion of face in their theory of politeness forms in language [4]. They characterized the degree of face threat of a given speech act as a function of power, social distance, and the intrinsic threat (imposition) imposed by the speech act.

Based on our own analysis of the use of social dialogue within task interactions (such as service encounters), we have further extended Brown and Levinson's model for determining face threats. Given the relational model presented above, the introduction of conversational topics which are at a significantly "deeper" level of familiarity than is expected relative to the existent relationship and activity are seen as a face threat. For example, if a stranger on the street approached you and asked you how much money you had in your bank account (or even what kind of toothpaste you used), you would likely perceive this as an intrusion and a threat to your face.

How can speakers change these dimensions of trust? An obvious strategy for effecting changes to the familiarity dimension of the relationship model is for the speaker to disclose information about him/herself and induce the listener to do the same. Social penetration theory has much to say about the self-disclosure process and its effect on not only the familiarity dimension of the relational model, but the affect dimension as well. There is a strong correlation between self-disclosure and liking (we like people who engage in more intimate disclosures, and we tend to disclose more to people we like). In addition, the principle of self-disclosure reciprocity states that one interlocutor's disclosure is likely to elicit from the other disclosures matched in topical content and depth [2]. Another way of changing the dimensions of trust in conversation is to engage in small talk.

Small Talk: Putting Trust-Elicitation into Practice

It's commonly thought that small talk is what strangers do when they must share a space for a time, but in general it can be taken as any talk in which interpersonal goals are emphasized and task goals are either non-existent or de-emphasized. Within task-oriented encounters, small talk can help humans or agents to achieve their 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 [16]. Small talk can also serve an exploratory function by providing a conventional mechanism for people to establish the capabilities and credentials ("communal common ground" [12]) of another human or a computational system. Small talk can build solidarity if the conversation involves a ritual of showing agreement with and appreciation of the conversational partner's utterances [20], [11, 27]. Finally, people and agents can use small talk to establish expertise, by relating stories of past successful problem-solving behavior, and to obtain information about the other that can be used indirectly to help achieve task goals (e.g., that the user drives a minivan increases the probability that the person has children).

Small talk can be used to address the face needs of interlocutors. In small talk, interlocutors take turns showing agreement with and appreciation of the contributions of the speaker, and in so doing enhance each other's face [11, 27]. This builds solidarity among the interlocutors by demonstrating their "like mindedness". Of course, small talk can also be used in social situations as a prelude to other, more personal kinds of talk (such as "getting acquainted talk" [29]), once the interlocutors decide that they want to move on to the next stage of their relationship.

Small talk can also be used to address interlocutor's face, by defusing awkward silences between strangers, such as in waiting rooms or airplanes [20, 27]. This is more of a defensive use of small talk, in which the interlocutors are attempting to establish only a minimal level of solidarity.

Thus, small talk implements the conversational strategies listed above in order to build trust (see Figure 1). It acts on a peer relationship among interlocutors, and thus may help to side-step any power imbalance between them. It allows them to establish common ground and thereby increase their familiarity with each other. It increases solidarity

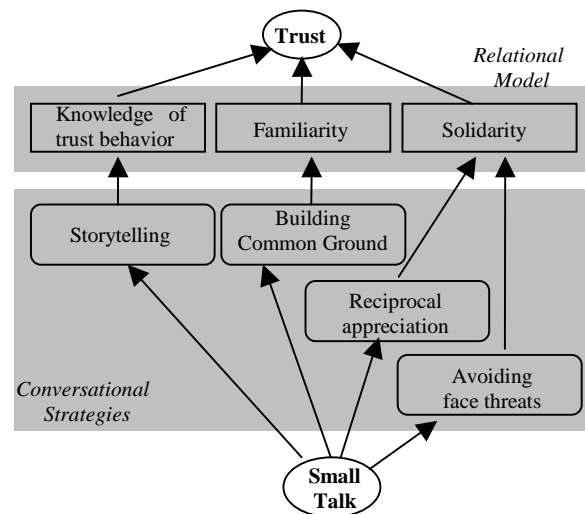


Figure 1. Influence of Small Talk on Trust

through taking turns showing appreciation and agreement with the contributions of one's partner. In fact, 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 [2]. 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.

AN IMPLEMENTATION: TRUST ELICITATION THROUGH SMALL TALK IN REA

REA is a real-time, multimodal, life-sized ECA, and her design is based on the FMTB model [7, 8]. REA has a fully articulated graphical body, can sense the user passively through cameras and audio input, and is capable of speech with intonation, facial display, and hand gesture. REA is displayed on a large projection screen, in front of which the user stands (see Figure 2). Two cameras mounted on top of the screen track the user's head and hand positions, while a microphone captures speech input. A single SGI Octane computer runs the graphics and conversation engine of Rea, while several other computers manage the speech recognition and generation, and image processing.

Rea simultaneously processes the organization of conversation and its content. When the user makes cues typically associated with turn taking behavior such as gesturing, Rea allows herself to be interrupted, and then takes the turn again when she is able. She is able to initiate conversational repair when she misunderstands what the user says, and can generate combined voice and gestural output. An incremental natural language generation engine based on [28], and extended to synthesize redundant and complementary conversational hand gestures, generates Rea's responses.

REA is an acronym for "Real Estate Agent", and within this domain we are currently focused on modeling the initial interview with a prospective buyer. Real estate sales was selected specifically for the opportunity to explore a task domain in which a significant amount of social dialogue normally occurs.

Implementing Relational Strategies in REA

Within initial interactions between professionals and their clients, small talk is often used to build trust and solidarity. This is especially important in real estate sales, where there is both a very significant commitment at stake and the buyer-agent relationship must continue for several weeks or months until a transaction is closed.

For the purpose of trust elicitation and small talk, we have constructed a new kind of discourse planner that can interleave small talk and task talk during the initial buyer interview, based on the model outlined above. Given that many of the goals in a relational conversational strategy are non-discrete (e.g., minimize face threat), and that trade-offs among multiple goals have to be achieved at any given time, we have moved away from static world discourse planning, and are using an activation network-based



Figure 2. User interacting with Rea

approach based on Maes' *Do the Right Thing* architecture [19]. This architecture provides the capability to transition smoothly from deliberative, planned behavior to opportunistic, reactive behavior, and is able to pursue multiple, non-discrete goals. In our implementation each node in the network represents a conversational move that REA can make.

During task talk, REA asks questions about users' buying preferences, such as the number of bedrooms they need. During small talk, REA can talk about the weather, events and objects in her shared physical context with the user (e.g., the lab setting), or she can tell stories about the lab, herself, or real estate.

REA's contributions to the conversation are planned in order to minimize the face threat to the user, and maximize trust, while pursuing her task goals in the most efficient manner possible. That is, Rea attempts to determine the face threat of her next conversational move, assesses the solidarity and familiarity which she currently holds with the user, and judges which topics will seem most relevant and least intrusive to users. As a function of these factors, Rea chooses whether or not to engage in small talk, and what kind of small talk to choose. The selection of which move should be pursued by REA at any given time is thus a non-discrete function of the following factors:

- Closeness -- Rea continually assesses her "interpersonal" closeness with the user, which is a composite representing depth of familiarity and solidarity, modeled as a scalar quantity. Each conversational topic has a pre-defined, pre-requisite closeness 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 finance.
- Topic -- Rea keeps track of the current and past conversational topics. Conversational moves which stay within topic (maintain topic coherence) are given preference over those which do not. In addition, Rea can plan to execute a sequence of moves 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).
- Relevance -- Rea maintains a list of topics that she thinks the user knows about, and the discourse planner prefers

moves which involve topics in this list. The list is initialized to things that anyone talking to Rea would know about--such as the weather outside, Cambridge, MIT, or the laboratory that Rea lives in.

- Task goals -- Rea has a list of prioritized goals to find out about the user's housing needs in the initial interview. Conversational moves which directly work towards satisfying these goals (such as asking interview questions) are preferred.
- Logical preconditions -- Conversational moves have logical preconditions (e.g., it makes no sense for Rea to ask users what their major is until she has established that they are students), and are not selected for execution until all of their preconditions are satisfied.

One advantage of the activation network approach is that by simply adjusting a few gains we can make REA more or less coherent, more or less polite (attentive to closeness constraints), more or less task-oriented, or more or less deliberative (vs. reactive) in her linguistic behavior.

In the current implementation, the dialogue is entirely REA-initiated, and user responses are recognized via a speaker-independent, grammar-based, continuous speech recognizer (currently IBM ViaVoice). The active grammar fragment is specified by the current conversational move, and for responses to many Rea small talk moves the content of the user's speech is ignored; only the fact that the person responded at all is enough to advance the dialogue.

At each step in the conversation in which Rea has the floor (as tracked by a conversational state machine in Rea's Reaction Module [9]), the discourse planner is consulted for the next conversational move to initiate. At this point, activation values are incrementally propagated through the network (following [19]) until a move is selected whose preconditions are satisfied and whose activation value is over a specified threshold.

Shifts between small talk moves and task moves are marked by conventional contextualization cues--discourse markers and beat gestures. Discourse markers include "so" on the first small talk to task talk transition, "anyway" on resumption of task talk from small talk, and "you know" on transition to small talk from task talk [12].

Within this framework, Rea decides to do small talk whenever closeness with the user needs to be increased (e.g., before a task query can be asked), or the topic needs to be moved little-by-little to a desired topic and small talk contributions exist which can facilitate this. The activation energy from the user relevance condition described above leads to Rea starting small talk with topics that are known to be in the shared environment with the user (e.g., talk about the weather or the lab).

Example Interactions

An interview between REA and a user typically proceeds as shown in the following dialogue. (User responses are only shown in positions in which they effect the selection of subsequent moves; Rea contributions are her actual utterances)

1. That microphone is terrible, I hate using those things.
2. Sorry about my voice, this is some engineer's idea of natural sounding.

3. Are you one of our sponsors? *User: Yes*
4. Were you at our last sponsor meetings?
5. I got so exhausted at the last sponsor meeting I think I was starting to lose my voice by the end.
6. So, where would you like to live?
7. How many bedrooms do you need?
8. Do you need access to the subway?
9. Is one bath enough?
10. You know, Boston is certainly more expensive than it used to be.
11. Anyway, what can you afford?
12. What kind of down payment can you make?
13. Let me see what I have available.

Dialogue 1. "Small Talk REA"

In this example, REA opens with small talk moves regarding things in her shared physical environment with the user (1-2). She then proceeds to small talk related to sponsors (after establishing that the user is a sponsor). After a few turns, enough closeness has been established (simply by doing small talk) that REA can move into task talk (6-9). However, before bringing up the topic of finance--a topic that is potentially very face threatening for the user--REA decides that additional closeness needs to be established, and moves back into small talk (10). This small talk move not only increases closeness but shifts the topic to finance, enabling REA to then bring up the issue of how much the user is able to afford (11-12).

If REA's adherence to closeness preconditions is reduced, by decreasing the contributions of these preconditions to the activation of conversational moves, this results in her engaging in less small talk and being more task goal oriented. If everything else is held constant (relative to the prior example) the following dialogue is produced.

1. So, where would you like to live?
2. What can you afford?
3. What kind of down payment can you make?
4. How many bedrooms do you need?
5. Do you need access to the subway?
6. Is one bath enough?
7. Let me see what I have available.

Dialogue 2. "Task-only REA"

In this example, REA does not perform any small talk and sequences the task questions in strictly decreasing order of priority.

EVALUATION

A formal model to relate trust and the use of small talk is of interest to the Communications community. The implementation of a small talk planning engine is of interest to the AI and Computational Linguistics communities. But, does small talk produced by an ECA in a sales encounter have any effect whatsoever on computer-human interaction? In order to evaluate whether an ECA's social dialogue can actually build trust and solidarity with users, we conducted an empirical study in which subjects were interviewed by Rea about their housing needs, shown two "virtual" apartments, and then asked to submit a bid on one of them. For the purpose of the experiment, Rea was

controlled by a human wizard and followed scripts identical to the output of the planner (but faster, and not dependent on automatic speech recognition or computational vision). Users interacted with one of two versions of Rea which were identical except that one had only task-oriented dialogue (TASK condition) while the other also included the social dialogue designed to avoid face threat, and increase trust (SOCIAL condition).

Our hypotheses follow from the literature on small talk and on trust among humans. We expected subjects in the SOCIAL condition to trust Rea more, feel closer to Rea, like her more, and feel that they understand each other more than in the TASK condition. We also expected users to think the interaction was more natural, lifelike, and comfortable in the SOCIAL condition. Finally, we expected users to be willing to pay Rea more for an apartment in the SOCIAL condition, given the hypothesized increase in trust.

Experimental Methods

Subjects. 18 people participated in the experiment (61% male and 39% female). Subjects were primarily students, and were recruited through ads on several college campuses.

Apparatus. An experiment room was constructed with one entire wall as a rear-projection screen, allowing Rea to appear life-sized on the screen, in front of the 3D virtual apartments she showed. Rea's synthetic voice was played through two speakers on the floor in front of the screen. Two video cameras and an omnidirectional microphone enabled recording of the subject's verbal and nonverbal behavior during the experiment.

The wizard sat behind the rear projection screen and controlled Rea's responses and sequencing through the interaction script via a computer. The script included verbal and nonverbal behavior specifications for Rea (e.g., gesture and gaze commands as well as speech), and embedded commands describing when different rooms in the virtual apartments should be shown. Three pieces of information obtained from the user during the interview were entered into the control system by the wizard: the city the subject wanted to live in; the number of bedrooms s/he wanted; and how much s/he were willing to spend. The first apartment shown was in the specified city, but had twice as many bedrooms as the subject requested and cost twice as much as s/he could afford (they were also told the price was "firm"). The second apartment shown was in the specified city, had the exact number of bedrooms requested, but cost 50% more than the subject could afford (but this time, the subject was told that the price was "negotiable"). The scripts for the TASK and SOCIAL condition were identical, except that the SOCIAL script had additional small talk utterances added to it, similar to those shown in Dialogue 1, above. The part of the script governing the dialogue from the showing of the second apartment through the end of the interaction was identical in both conditions.

Procedure. Subjects were told that they would be interacting with Rea, who played the role of a real estate agent and could show them apartments she had for rent. They were told that they were to play the role of someone looking for an apartment in the Boston area, and that they were to stand in front of Rea and talk to her "just like you would to another person".

Subjects were then shown a brief (one minute) video of Rea on a small monitor, giving additional instructions regarding her speech recognition software. The purpose of this was to both reduce the "novelty effect" when Rea first appeared on the big projection screen, and to ensure the deception (use of a wizard) was effective. Subjects then interacted with Rea, after which they were asked to fill out a questionnaire.

Manipulation check. Three questions concerning the amount of small talk used by Rea were included on the questionnaire, both for development feedback and for manipulation checks. That is, subjects were asked, for example, how quickly Rea got down to business. If there is a perceivable difference between the small talk and task-only conditions, then subjects should believe that task-only Rea got down to business more quickly. In addition, during the debriefing session, subjects were asked about the amount of small talk Rea did. All subjects in the TASK condition mentioned that they did not engage in any small talk, while all subjects in the SOCIAL condition commented on some aspect of the small talk Rea performed. There was also a significant difference ($F=5.9$; $p < .03$) such that users believed that Rea got down to business more quickly in the task-only condition than in the small talk condition.

Measures.

Trust was measured by a standardized trust scale [31] ($\alpha = .88$ [22]).

Liking of Rea, Closeness to Rea, Warmth of Rea, Naturalness of the Interaction, and Enjoyment of the Interaction were measured by single items on nine-point Likert scales.

Amount Willing to Pay was computed as follows. During the interview, Rea asked subjects how much they were able to pay for an apartment; subjects' responses were entered as \$X per month. Rea then offered the second apartment for \$Y (where $Y = 1.5 X$), and mentioned that the price was negotiable. On the questionnaire, subjects were asked how much they would be willing to pay for the second apartment, and this was encoded as Z. The task measure used was $(Z - X) / (Y - X)$, which varies from 0% if the user did not budge from their original requested price, to 100% if they offered the full asking price.

Given results in the literature on the relationship between user personality and preference for computer behavior, we were concerned that subjects might respond differentially to social dialogue based on predisposition. Thus, we also included composite measures for introversion and extroversion on the questionnaire.

Extrovertedness was an index composed of seven Wiggins [32] extrovert adjective items: Cheerful, Enthusiastic, Extroverted, Jovial, Outgoing, and Perky. It was used for assessment of the subject (Cronbach's alpha = .94 [22]).

Introvertedness was an index composed of seven Wiggins [32] introvert adjective items: Bashful, Introverted, Inward, Shy, Undemonstrative, Unrevealing, and Unsparkling. It was used for assessment of the subject (alpha = .83 [22]).

Finally, observation of the videotaped data made it clear that some subjects took the initiative in the conversation, while others allowed Rea to lead. Unfortunately, Rea is not yet able to deal with user-initiated talk, and so user initiative often led to Rea interrupting the speaker. To assess the effect of this phenomenon, we therefore divided subjects into *passive* (below the mean on number of user-initiated utterances) and *initiators* (above the mean on number of user-initiated utterances). To our surprise, these measures turned out to be independent of introversion/extroversion, and to not be predicted by these latter variables.

Results

The most striking results were interactions between, on the one hand, intro/extroversion and trust and, on the other hand, initiative/passivity and the composite of engagement, and the single variables of interest, naturalness, and degree to which subjects felt Rea knew them.

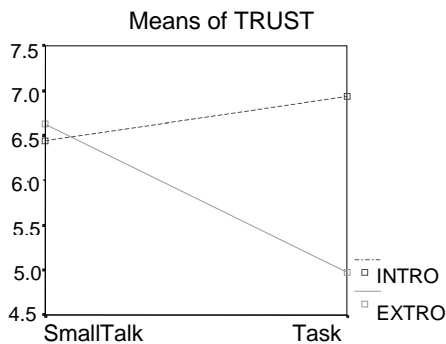


Figure 3: Trust Estimation by introverts & extroverts

Figure 3 shows the interaction between intro/extroversion and trust ($F=4.5$; $p<.05$). These results indicate that small talk had essentially no effect on the trust assessment of Introverts. However, this kind of social dialogue had a significant effect on the trust assessment of extroverts, in fact social dialogue seemed to be a pre-requisite for establishing the same level of trust for extroverts as that experienced by introverts.

Figure 4 shows the interaction between initiator/passivity and engagement. These results indicate that active users felt more engaged with Rea using small talk, while passive users felt more engaged with task-only dialogue ($F=9.7$; $p < .01$).

Likewise, more active users felt as if the interaction were more natural ($F=4.7$; $p < .05$); more interesting ($F=5.2$; $p <$

$.05$), and as if Rea came to know them better ($F=6.8$; $p < .03$) when she used small talk.

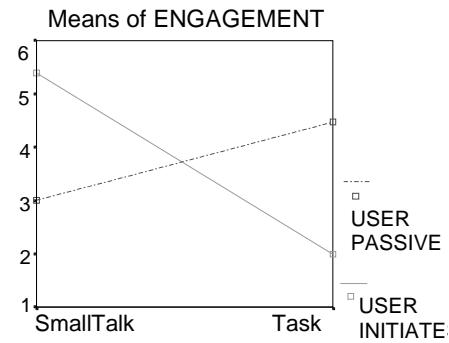


Figure 4: Engagement by initiators vs. passive speakers

In all of these cases, users who reach out more towards other people are more susceptible to relationship building. And, those people need some relational conversational strategies in order to trust the interface.

No significant effects were found on Amount Willing to Pay across conditions. Although we had assumed that there would be a strong correlation between trust in Rea and this measure, there may be other factors involved in the pricing decision, and we plan to investigate these in the future.

CONCLUSION

Relational intelligence includes knowledge of when and how to use language to achieve social goals. This knowledge is crucial for our computational agents if they are to be as effective as people, and if we want people to be able to use our agents easily, efficiently, and cooperatively.

As embodied conversational agents become ubiquitous, the ability for them to establish and maintain social relationships with us will become increasingly important.

We are currently investigating the implementation of other forms of social dialogue and additional relational strategies, as well as expanding the dyadic relationship model used in our discourse planner.

For the moment, however, we have shown that models of social dialogue can be formalized in an implementable way, and that their evaluation demonstrates the importance of the phenomenon to a well-defined subset of users. The study of human-computer relationships is a new field which exists at the nexus of research into human-computer interaction, human social psychology, sociology, and linguistics. The study of how to constitute relationships through language will inform our growing ability to emulate aspects of humans in the service of efficient interaction between humans and machines.

REFERENCES

1. Andre, E., Muller, J., and Rist, T., "The PPP Persona: A Multipurpose Animated Presentation Agent," presented at Advanced Visual Interfaces, 1996.

2. Berscheid, E. and Reis, H., "Attraction and Close Relationships," in *The Handbook of Social Psychology*, D. Gilbert, S. Fiske, and G. Lindzey, Eds. New York: McGraw-Hill, 1998.
3. Beskow, J. and McGlashan, S., "Olga: a conversational agent with gestures," presented at IJCAI 97, 1997.
4. Brown, P. and Levinson, S., "Universals in language usage: Politeness phenomena," in *Questions and Politeness: Strategies in Social Interaction*, E. Goody, Ed. Cambridge: Cambridge University Press, 1978.
5. Brown, R. and Gilman, A., "The pronouns of power and solidarity," in *Language and Social Context*, P. Giglioli, Ed. Harmondsworth: Penguin, 1972, pp. 252-282.
6. Cassell, J. and Bickmore, T., "External Manifestations of Trustworthiness in the Interface," *Communications of the ACM*, vol. to appear, 2000.
7. Cassell, J., Bickmore, T., Billinghurst, M., Campbell, L., Chang, K., Vilhjalmsson, H., and Yan, H., "Embodiment in Conversational Interfaces: Rea," presented at CHI 99, Pittsburgh, PA, 1999.
8. Cassell, J., Bickmore, T., Campbell, L., Vilhjalmsson, H., and Yan, H., "Human Conversation as a System Framework: Designing Embodied Conversational Agents," in *Embodied Conversational Agents*, J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, Eds. Cambridge, MA: MIT Press, 2000.
9. Cassell, J., Bickmore, T., Vilhjalmsson, H., and Yan, H., "More Than Just a Pretty Face: Affordances of Embodiment," presented at IUI 2000, New Orleans, Louisiana, 2000.
10. Cassell, J., Sullivan, J., Prevost, S., and Churchill, E., *Embodied Conversational Agents*. Cambridge: MIT Press, 2000.
11. Cheepen, C., *The Predictability of Informal Conversation*. New York: Pinter, 1988.
12. Clark, H. H., *Using Language*. Cambridge: Cambridge University Press, 1996.
13. Dehn, D. M. and Mulken, S. v., "The Impact of Animated Interface Agents: A Review of Empirical Research," University of Saarland, Saarbrücken, Germany 1999.
14. Fogg, B. J. and Tseng, H., "The Elements of Computer Credibility," presented at CHI 99, 1999.
15. Goffman, I., "On face-work," in *Interaction Ritual: Essays on Face-to-Face Behavior*. New York: Pantheon, 1967, pp. 5-46.
16. Jaworski, A. and Coupland, N., *The Discourse Reader*. London: Routledge, 1999.
17. Lester, J., Stone, B., and Stelling, G., "Lifelike Pedagogical agents for Mixed-Initiative Problem Solving in Constructivist Learning Environments," *User Modeling and User-Adapted Interaction*, vol. 9, pp. 1-44, 1999.
18. Lester, J. C., Voerman, J. L., Towns, S. G., and Callaway, C. B., "Cosmo: A Life-like Animated Pedagogical Agent with Deictic Believability," presented at IJCAI 97, 1997.
19. Maes, P., "How to do the right thing," *Connection Science Journal*, vol. 1, 1989.
20. Malinowski, B., "The problem of meaning in primitive languages," in *The Meaning of Meaning*, C. K. Ogden and I. A. Richards, Eds.: Routledge & Kegan Paul, 1923.
21. Moon, Y., "Intimate self-disclosure exchanges: Using computers to build reciprocal relationships with consumers," Harvard Business School, Cambridge, MA Working paper 99-059, 1998.
22. Nass, C. and Lee, K., "Does Computer-Generated Speech Manifest Personality? An Experimental Test of Similarity-Attraction," presented at CHI 2000, The Hague, Amsterdam, 2000.
23. Reeves, B. and Nass, C., *The Media Equation: how people treat computers, televisions and new media like real people and places*. Cambridge: Cambridge University Press, 1996.
24. Resnick, P. V. and Lammers, H. B., "The Influence of Self-esteem on Cognitive Responses to Machine-Like Versus Human-Like Computer Feedback," *The Journal of Social Psychology*, vol. 125, pp. 761--769, 1985.
25. Rickel, J. and Johnson, W. L., "Animated agents for procedural training in virtual reality: Perception, cognition, and motor control.," *Applied Artificial Intelligence*, 1998.
26. Rickenberg, R. and Reeves, B., "The Effects of Animated Characters on Anxiety, Task Performance, and Evaluations of User Interfaces," presented at CHI 2000, The Hague, Amsterdam, 2000.
27. Schneider, K. P., *Small Talk: Analysing Phatic Discourse*. Marburg: Hitzeroth, 1988.
28. Stone, M. and Doran, C., "Sentence Planning as Description Using Tree-Adjoining Grammar," presented at ACL, 1997.
29. Svennevig, J., *Getting Acquainted in Conversation*. Philadelphia: John Benjamins, 1999.
30. Thorisson, K. R., "Gandalf: An Embodied Humanoid Capable of Real-Time Multimodal Dialogue with People," presented at Autonomous Agents '97, 1997.
31. Wheelless, L. and Grotz, J., "The Measurement of Trust and Its Relationship to Self-Disclosure," *Human Communication Research*, vol. 3, pp. 250-257, 1977.
32. Wiggins, J., "A psychological taxonomy of trait-descriptive terms," *Journal of Personality and Social Psychology*, vol. 37, pp. 395-412, 1979.