(in press) User Modeling and Adaptive Interfaces

Negotiated Collusion: Modeling Social Language and its Relationship Effects in Intelligent Agents

Justine Cassell, Timothy Bickmore

MIT Media Lab
20 Ames St., E15-315
Cambridge, MA 02139 USA
+1 617 253 4899
{justine, bickmore}@media.mit.edu

"This evidence leads us to wonder whether intimacy is as much a '*negotiated collusion*' as it is a state of 'true oneness'"
(Brown & Rogers, 1991)

Abstract

Building a collaborative trusting relationship with users is crucial in a wide range of applications, such as advice-giving or financial transactions, and some minimal degree of cooperativeness is required in all applications to even initiate and maintain an interaction with a user. Despite the importance of this aspect of human-human relationships, few intelligent systems have tried to build user models of trust, credibility, or other similar interpersonal variables, or to influence these variables during interaction with users. Humans use a variety of kinds of social language, including small talk, to establish collaborative trusting interpersonal relationships. We argue that such strategies can also be used by intelligent agents, and that embodied conversational agents are ideally suited for this task given the myriad multimodal cues available to them for managing conversation. In this article we describe a formal theory of the relationship between social language and interpersonal relationships, a new kind of discourse planner that is capable of generating social language to achieve interpersonal goals, and an actual implementation in an embodied conversational agent. We discuss an evaluation of our system in which the use of social language was demonstrated to have a significant effect on users' perceptions of the agent's knowledgableness and ability to engage users, and on their trust, credibility, and how well they felt the system knew them, for users manifesting particular personality traits.

KEYWORDS: Embodied conversational agent, small talk, trust, social interface, dialogue

1 Introduction

In this article we address a new aspect of user modeling – assessing the psychosocial relationship between the person and the computer. And we introduce new methods for adapting the computer's behavior to the user model, as well as for explicitly and dynamically changing this relationship through the use of social talk. Human-human dialogue does not just comprise statements about the task at hand, about the joint and separate goals of the interlocutors, and about their plans. In human-human conversation participants often engage in talk that, on the surface, does not seem to move the dialogue forward at all. However, this talk – about the weather, current events, and many other topics without significant overt relationship to the task at hand -- may, in fact, be essential to how humans obtain information about one another's goals and plans and decide whether collaborative work is worth engaging in at all. For example, realtors use small talk to gather information to form stereotypes (in Rich's sense (Rich, 1979) of a collection of frequently co-occurring characteristics); of their clients – people who drive minivans are more likely to have children, and therefore to be searching for larger homes in neighborhoods with good schools. Realtors also use small talk to increase intimacy with their clients, to establish their own expertise, and to manage how and when they present information to the client. In this article we discuss the implementation and evaluation of an embodied conversational agent that can engage in small talk of this same sort, and use that small talk as a way to evoke collaborative behaviors in users. We argue that the user model – commonly thought to comprise a model of the human's goals, plans and knowledge – should also cover the user's judgment of the system's reliability, collaborativeness and trustworthiness. In our evaluation we discover that small talk can contribute positively to users' perceptions of a system, but that it has a differential effect on users, depending on their personality profiles. We end by discussing how user modeling could be extended not just to the short term social features of closeness and trust, but also to modeling personality profiles or stereotypes to improve interaction between humans and computers.

We begin by illustrating the phenomenon of interest with an example from an actual conversation between a realtor and a client.

2 An Example from Human Interaction

The following is an excerpt from an actual interview between a realtor (R): and two clients (C1 and C2)

1. R: Alright. [From your names] I can see that you're related. By marriage, right?
2. C1: Recently.
3. C2: Newlyweds.
4. R: Really? When?
5. C1: June eleventh.
6. R: Congratulations.
7. C1: We're also expecting a baby.
8. R: Holy cow.
9. C1: In May. So. And we want to buy a house.
10. R: You guys don't fool around, do you? You took awhile to decide then it was like, let's do it all.
11. C1: Moving into the new Millenium. So.
12. R: That's the way. Do you know if it's going to be a boy or a girl?
13. C1: I find out next week.
14. R: When's the due date?
15. C1: May 29th.
16. R: Very good.
17. C1: Yea. Good timing.
18. R: Awesome. You want the house before the child?

The clients reply to a simple question about their common last name by disclosing unrequested information – that they are recently married. Rather than bringing the conversation back to their housing needs, the realtor continues to foster this interpersonal aspect of the conversation by asking for the marriage date. But she manages to encourage the social chitchat in a direction which leads to essential information – how quickly the clients are willing to purchase their new home. We can imagine that this realtor has constructed a model of her clients that represents a number of features relevant to her collaboration with them. These features, we claim, are both directly related to the task at hand, and related to the nature of the relationship between the participants. That is, the realtor is modeling a goal-oriented feature of the clients -- the size of their family, and therefore how big a house they will need – and a social feature of the clients – how close they feel with her at that moment, and therefore how likely they are to want to work with her. It is this second aspect of user modeling that is the topic of the present article.

People 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. Our realtor is making it clear to her clients that she cares about their lives (and, by extension, will be on their side during the purchase of a home).

Computer interface agents may also profitably use social 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, facilitates cooperation. But cooperativeness is a relationship that implies the perception that each of two entities has of the other, and this means that the user is also simultaneously constructing and maintaining a model of the system, and it is this model of the system which determines in large part the user's actions. How must the system act, not just to construct a user model, but also in order to evoke a model that facilitates trust & collaboration on the part of the user? We argue that the system may adopt some of the same strategies as used by humans: increasing intimacy over the course of a conversation, decreasing interpersonal distance, using non-explicit ways of achieving conversational goals, displaying expertise, managing "face" by leading up to face-threatening topics slowly. Further, we believe that these strategies may be realized linguistically by agents in the some of the same ways they are realized by humans, by small talk, for example. This argument rests on the assumption that a human-computer interface based on familiar human social rules and conventions will be easy for people to engage with, and successful in evoking familiar human responses. We will not directly test this assumption, however we do evaluate the success of the interface that rests on these principles, and find our assumption well-warranted.

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 (Cassell, Sullivan, Prevost, & Churchill, 2000). These nonverbal channels are also especially crucial for the management of the 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 interpersonal activities.

3 Related Work

In this article we are interested in modeling the relationship that humans maintain with a computer, and methods for affecting that relationship through particular dialogue strategies. Very little work has been done in this area of modeling dialogue and social relationships. We first motivate our work by looking at what is known about the social nature of users' relationships with their computers, and how embodiment

may play a role. Given the dearth of research in this area, we then step back and review the social psychology and sociolinguistic literature on interpersonal relationships and conversation. We extend and formalize these previous models of how conversation plays a role in interpersonal relationships, and then we present our own model of social language between humans and computers. Later in the paper, after this model is presented, we discuss how it has been implemented in an actual embodied conversational agent, and how it has been evaluated in human-computer interaction.

3.1 Interpersonal Relationships with Agents

In a series of studies, researchers in the "Computers As Social Actors" paradigm 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 (Reeves & Nass, 1996). Morkes, Kernal and Nass demonstrated that computer agents that use humor are rated as more likable, competent and cooperative than those that do not (Morkes, Kernal, & Nass, 1998). 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 (Moon, 1998). In a different paradigm, Mark and Becker (Mark & Becker, 1999) studied how social conventions could affect interpersonal relationships between humans meeting in cyberspace. These are examples of persuasion tactics (Fogg, 1999) employed by computers to change the beliefs, feelings, thoughts of users.

Of course the social influence strategies of 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 (Reeves & Nass, 1996) (Nass & Lee, 2000). 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) (Resnick & Lammers, 1985). 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 (Rickenberg & Reeves, 2000).

Few of the results from these studies, however, have yet found their way into the user modeling literature, or been implemented in working systems. Castelfranchi and de Rosis (Castelfranchi & de Rosis, 1999) describe the conditions under which a computational system might wish to deceive. Pautler (Pautler, 1998) used a unidimensional relationship model to represent the relationship between users and their colleagues in terms of perlocutionary acts. Ardissono et. al (Ardissono, Boella, & Lesmo, 1999) describe a formal framework for generating indirect speech acts as a function of face threat and face saving. Ward (Ward, 1997) has begun to model 'real-time social skills' – that is, the ability of one person to sense the fleeting changes in another's mood, desires, intentions. His model depends on analyzing the prosody of a user's utterances, from it inferring fleeting changes of state, and then responding with different kinds of non-verbal acknowledgment feedback (for example "uh-huh" vs. "hmmm"). Along the personality dimension, Ball and Breese (Ball & Breese, 2000) have experimented with Bayesian networks to model the emotion and personality of a user, and then to choose corresponding responses on the part of a system.

3.2 Embodied Conversational Agents

User modeling includes recognizing some aspect of the user, and then constructing appropriate responses. As Ward (op. cit) has pointed out, in order to model the user's social states, the system needs to be able to recognize embodied behaviors such as gestures and non-speech sounds, and in order to adapt to the user,

the system must be able to engage in similarly embodied actions. Work on the development of ECAs, as a distinct field of development, is best summarized in (Cassell, Sullivan, Prevost, & Churchill, 2000). In addition to REA (Cassell, et al., 1999) (described below), some of the other major ECA systems developed to date are Steve (Rickel & Johnson, 1998), the DFKI Persona (Andre, Muller, & Rist, 1996), Olga (Beskow & McGlashan, 1997), Gandalf (Thorisson, 1997), and pedagogical agents developed by Lester, et al, (Lester, Voerman, Towns, & Callaway, 1999). 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 such 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 (Dehn & Mulken, 1999), specifically examining evaluations of recent animated interface agents, conclude that the benefits of these systems are still 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 language for building user trust and diminishing interpersonal distance, and that is implemented in a domain in which exactly these abilities are key.

3.3 Dimensions of Interpersonal Relations in Conversation

One of the issues that arises when looking at how to recognize user feelings about a computer, and trying to determine how to influence them, is that interpersonal relationships can be measured and represented along many dimensions, including intimacy, solidarity, closeness, familiarity, and affiliation (Spencer-Oatey, 1996). Here we are interested in dimensions that have an effect on collaborative activity and trust and that can be employed to formulate a communicative strategy and so we base our user-computer social linguistic model on the dimensions of the 'interpersonal relations in conversation' model developed by Svennevig (Svennevig, 1999), which addresses directly the interaction between language and relationships. In what follows, we describe these four dimensions, and some strategies for affecting them, from Svennevig's own model, and then we lay out our own extensions to the model.

The first dimension of Svennevig's relational model is labeled *familiarity*. Based on social penetration theory (Berscheid & Reis, 1998), which claims to account for the establishment and growth of interpersonal relationships, this dimension describes the way in which relationships develop through the reciprocal exchange of information, beginning with relatively non-intimate topics and gradually progressing to more personal and private topics. The growth of a relationship can be represented in both the breadth (number of topics) and depth (public to private) of information disclosed.

Two other dimensions of Svennevig's relational model – *power* and *solidarity* – are based on work both in social psychology, and in linguistics that accounts for the usage of different forms of address (T-forms vs. V-forms for example (Brown & Gilman, 1972)). Power is the ability of 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, gender, etc.), and is very similar to the notion of social distance used by Brown and Levinson in their theory of politeness (Brown & Levinson, 1978). There is a correlation between frequency of contact and solidarity, but it is not necessarily a causal relation (Brown & Levinson, 1978; Brown & Gilman, 1972).

The fourth and final dimension of Svennevig's model is *affect*. This represents the degree of liking the interactants have for each other, and there is evidence that this is an independent relational attribute from

the above three (Brown & Gilman, 1989). In Pautler's computational model of social perlocutions, affect is the only dimension of relationship modeled (Pautler, 1998).

Although trust is also an essential part of human social relationships, and is often established through linguistic means, Svennevig does not include trust as one of the dimensions, since he believes it can be better viewed as a function or outcome of the above attributes, and not a dimension to be modeled independently. From other sources, we define trust 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," and one model of the development of trust 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" (Berscheid & Reis, 1998). In addition, disclosing information to another communicates that we trust that person to respond appropriately. Thus, trust is predicated on solidarity and familiarity, but also includes information about specific trusting behaviors, in addition to disclosure. The establishment of trust is crucial for human-computer interaction, since it is prerequisite to cooperative behavior on the part of the user, but we believe that a cognitive state of trust can be evoked in users by varying the dimensions of the social linguistic model (Cassell & Bickmore, 2000). Note that this formulation differs from recent work on trust in the computational community (Fogg & Tseng, 1999) in that work on trust in e-commerce or among agents often relies on transaction characteristics rather than interpersonal characteristics.

3.4 Strategies for Establishing Interpersonal Relations

People have myriad strategies available to them in conversation for establishing and maintaining the four dimensions of interpersonal relationships. Here we introduce three broad categories of interpersonal work, or strategies, that have shown to be effective in establishing and maintaining interpersonal relationships, and that are amenable to formal modeling: facework, establishing common ground, and affective strategies. First we describe the strategies and then we will turn to the kinds of talk that can realize them.

3.4.1 Facework

In Goffman's approach to social interaction, which set the base for future work, 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 (Goffman, 1967). The notion of "face" 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. 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, both for ourselves and for those we interact with, 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 (Brown & Levinson, 1978). They defined positive face as an individual's desire to be held in esteem by his/her interactants, and negative face as an individual's desire for autonomy, and 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. That is, the face threat to the hearer can be given by:

face threat = f($SA_{intrinsic}$, Power, Distance)
$SA_{intrinsic}$ = f(SA)
Where,
$SA_{intrinsic}$ = the intrinsic threat of the speech act
SA = denotes a class of speech acts
Power = power relationship between speaker and hearer
Distance = social distance between speaker and hearer

Figure 1: Brown & Levinson's Face Threat

The ‘intrinsic threat’ parameter accounts for the fact that certain speech acts are more of a threat than others. For example, an informing is less of a threat than a request for information which is less of a threat than a rejection. Distance is defined to be "a symmetric social dimension of similarity/difference within which the speaker and hearer stand for the purposes of this act", and is thus very similar to the notion of solidarity defined above. Power is identical to the definition given above.

If a significant threat will result from the speaker producing the indicated speech act, then the speaker has several options: 1) don't do the act; 2) do the act "off record"; 3) do the act "on record" with redressive action (negative politeness strategies); 4) do the act on record with redress action (positive politeness strategies); 5) do the act on record, "baldly". Following Grice’s (Grice, 1989) description of how to fail to fulfill the conversational maxims, these options are ranked in order of decreasing ability to mitigate a threat, thus the most threatening acts shouldn't be done at all, while the least threatening acts can be done baldly on record. Examples of "off record" acts are hinting and/or ensuring that the interpretation of the utterance is ambiguous (e.g., "I'm thirsty."). Negative politeness strategies include those which are oriented towards the autonomy concerns of the listener (e.g., "Could you bring me a drink?"), while positive politeness strategies address the esteem concerns of the listener (e.g., "Hey my friend, get me a drink.").

Svennevig, in turn, extended Brown and Levinson’s model by noticing that the threat perceived from different types of speech acts can change based on context, and in particular based on the relationship between the speaker and hearer (Svennevig, 1999). For example, close friends have established a set of mutual rights and obligations and thus do not experience certain acts (such as requests) as face threatening, but rather as confirming and reestablishing their relational bonds. Thus, his extension to the model can be characterized as:

face threat = f(SA, Power, Solidarity, Familiarity, Affect)

Where,

SA = denotes a class of speech acts (not mapped to an "intrinsic" threat value)

Figure 2: Svennevig's Face Threat

Key to our concerns here, where an agent will wish to recognize the user’s sense of distance, and actively decrease it, is Svennevig’s observation that politeness strategies can actually effect a change in interpersonal distance:

> The language forms used are seen as reflecting a certain type of relationship between the interlocutors. Cues may be used strategically so that they do not merely reflect, but actively define or redefine the relationship. The positive politeness strategies may thus … contribute to strengthening or developing the solidarity, familiarity and affective bonds between the interactants. The focus is here shifted from maintaining the relational equilibrium toward setting and changing the values on the distance parameter. (Svennevig, 1999); 46-47.

3.4.1.1 Our Extension to a Model of Face Threat

We have collected a number of task interaction dialogues (between realtors and clients, opticians and clients, and opticians and suppliers) and based on an analysis of the use of social dialogue within these task interactions, we have further extended Brown and Levinson's model for determining face threats. Given the relational model presented above, it is clear that the introduction of conversational topics which are at a significantly deeper level of familiarity than is expected relative to the existent relationship and activity will be seen as a face threat. For example, if a stranger on the street asked you how much money you had in your bank account, you would likely perceive this as a threat to your face. Such a kind of face

threat is key to task encounters where strangers must interact, and occasionally share personal information. We term this a "Social Penetration" threat, or SP_{threat}.

Topics that are at the appropriate level of familiarity but which "come out of the blue" also seem to us to be face threats, but have not been accounted for in a general way in previous theory. While a subset of these have been addressed in Brown and Levinson's theory (e.g., rejections), moves which are deemed dispreferred based solely on their sequential placement in conversation cannot be accounted for, given Brown & Levinson's use of isolated speech acts as their point of departure. Instances of such "sequential placement" threats are failing to demonstrate the relevance of a conversational story, appreciation of conversational stories following their conclusion (Jefferson, 1978), or introducing conversational topics or stories which are not related to the on-going discourse (not "locally occasioned" (Sacks, 1995)). Thus, for example, if you are telling your office mate a highly charged story about the crazy person in the subway that morning, and your office mate replies, not by nodding or otherwise acknowledging your story, but instead by asking if you want a cup of coffee, that will threaten your face. This variety of face threat must be dealt with in task-oriented dialogue of the sort engaged in by agents and users, in order to maintain the relevance of task-oriented and socially-oriented talk as the dialogue advances.

Our resulting model of face threat then becomes:

face threat $= \mathbf{f(SA_{threat}, SP_{threat})}$
$\mathbf{SA_{threat}} = \mathbf{f(SA_k, \{SA_1, \ldots, SA_j\}, Power, Solidarity, Familiarity, Affect)}$
$\mathbf{SP_{threat}} = \mathbf{f(FamiliarityDepth, TopicDepth)}$
Where,

- SA_{threat} = Threat due to the speech act.
- SP_{threat} = Threat due to violation of social penetration theory.
- SA_k = The class of speech act.
- $\{SA_1, \ldots, SA_j\}$ = The discourse context of speech acts into which SA_k will be introduced. For example, SA_1 could represent the overall conversation, and SA_j represents the activity which SA_k will become a constituent of.
- TopicDepth = The "depth" of the topic to be introduced (wrt social penetration theory).

Figure 3: Cassell & Bickmore's Face Threat

Facework is important primarily in preventing the dissolution of solidarity. That is, dispreferred and overly familiar conversational moves must be avoided in order to maintain solidarity at the level where it started. In the next section, we turn to a strategy for reducing interpersonal distance.

3.4.2 Establishing Common Ground

Personal information which is known by all interactants to be shared (mutual knowledge) is said to be in the "common ground" (Clark, 1996). The principle way for personal information to move into the common ground is via face-to-face communication, since all interactants can observe the recognition and acknowledgment that the information is in fact mutually shared. One strategy for effecting changes to the familiarity dimension of the relationship model is for speakers to disclose information about themselves – moving it into the common ground – and induce the listener to do the same. Another strategy is to talk about topics that are obviously in the common ground – such as the weather, physical surroundings, and other topics available in the immediate context of utterance.

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 (Berscheid & Reis, 1998). As described above, self-disclosure has been shown to play a significant role in human-computer interaction (Moon, 1998). We depend on the common ground aspect of disclosure in our work, by having our system begin an interaction by discussing topics that are clearly shared.

3.4.3 Coordination

The process of interacting with a user in a fluid and natural manner may increase the user's liking of the agent, and user’s positive affect, since the simple act of coordination with another appears to be deeply gratifying. "Friends are a major source of joy, partly because of the enjoyable things they do together, and the reason that they are enjoyable is perhaps the coordination." (Argyle, 1990). Studies of mother-infant interactions support the innate appeal of coordination, and the proposed link between synchrony and attachment (Depaulo & Friedman, 1998). A happy agent may also cause "emotional contagion" via motor mimicry, which has been shown to induce affective reactions (a smiling agent causes the user to smile which causes the user to feel happy) (Depaulo & Friedman, 1998). Buck terms the phenomenon a "conversation between limbic systems" (Buck, 1993). Thus, an agent that is able to closely sychronize its speech and nonverbal conversational behaviors is likely to increase the user’s positive affect towards it. This phenomenon is one of the reasons for embodying an agent, and providing it with a range of both verbal and nonverbal behaviors. Fundamentally, changes in interactants’ transient affect have a cumulative effect on their affective stance towards the relationship; users who are repeatedly made happy by an agent will likely increase their liking for the agent over time (this is the mechanism represented in Pautler's model (Pautler, 1998)).

3.5 Language for Carrying out Strategies for Establishing Interpersonal Relations

Thus far we have discussed the dimensions of interpersonal relations as they exist in conversation, and we have discussed strategies for establishing, maintaining, and changing those interpersonal relations. In this section we address the actual linguistic forms that realize those strategies. That is, we now turn to how particular kinds of talk can realize facework, common ground, and affective strategies. We will concentrate on *small talk* although other kinds of social language (such as gossip and jokes) are also important, and remain exciting topics for further research in this area.

It is commonly thought that small talk is what strangers do when they must share a small space for a large period of 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 (including social chit chat, conversational stories, asides). As illustrated above, 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 situation (Jaworski & Coupland, 1999). Small talk can also serve an exploratory function by providing a conventional mechanism for people to establish their capabilities and credentials. The realtor in the dialogue cited above, for example, later used to small talk to demonstrate her skills by telling a short anecdote about how she had sold a house to her very own tenant, and how successful that sale had been. Small talk can build solidarity if the conversation involves a ritual of showing agreement with and appreciation of the conversational partner’s utterances (Malinowski, 1923) (Schneider, 1988) (Cheepen, 1988). 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 client is pregnant increases the probability that the person will require a two-bedroom or larger home).

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 (Cheepen, 1988; Schneider, 1988). 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" (Svennevig, 1999)), 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 (Malinowski, 1923; Schneider, 1988). This is more of a defensive use of small talk, in which the interlocutors are attempting to establish only a minimal level of solidarity.

3.5.1 How Small Talk Works

The topics in small talk are highly constrained, and typically begin with subjects in the interlocutors' immediate shared context (e.g., the weather), since that is both safe and can be presumed to be in the common ground. Topics can then either progress out to the shared sociocultural context (e.g., economy, "light politics"), or in to personal topics of the participants. The former approach is more typically followed in social contexts (e.g., parties) while the latter is more typical of strangers who must address an awkward silence between them (Schneider, 1987).

When used to address positive face wants, interlocutors show increased attentiveness towards each other. Stylistically, then, small talk can be seen as a kind of ostensible communication (Clark, 1996) in which the interlocutors are pretending to be close friends or acquaintances, while keeping the discourse topics at a safe level of interpersonal distance. This being the case, interlocutors engaged in small talk show signs of positive affect in their speech, conveying some of the signs of "interpersonal warmth," including such behaviors as (Andersen & Guerrero, 1998):

- Proxemic behaviors: close conversational distance, direct body orientation, forward leans, communicating at the same level or in the same physical plane
- Oculesic behaviors: increased gaze, mutual eye contact, decreased eye movements
- Kinesic behaviors: smiling, general facial pleasantness, affirmative head nods, gestural animation, head tilts, bodily relaxation, lack of random movement, open body positions, postural congruence
- Vocalic behaviors: more variation in pitch, amplitude, duration and temp; reinforcing interjections such as "uh-huh" and "mm-hmmm"; greater fluency, warmth, pleasantness, expressiveness, and clarity; smooth turn-taking

Structurally, small talk has been characterized (Schneider, 1988) in terms of an initial question-response pair, followed by one of several types of third moves (echo question, check-back, acknowledgement, confirmation, evaluation), followed by zero or more synchronized "idling" moves. An example of such an exchange reported by Schneider is:

> A: It's a nice morning, isn't it?
> B: It's very pleasant.
> A: It is really, it's very pleasant, yes.
> B: Mhm.

Topic introduction also follows a number of structural constraints. Topics are negotiated among the interlocutors, rather than simply introduced by one speaker. The constraints on topic include the following (Svennevig, 1999):

- reportability - a presumption of interest by the interlocutors, established in negotiation. Topics can be proposed via "topical bids" ("I just got back from vacation.") and taken up via "topicalizers" ("Oh yea?") which indicate interest in the topic.

- projectability - a plan for the topic talk should be indicated in the proposal, usually by means of indicating the genre to be used -- narrative, argumentation, exposition, etc. ("oh, that reminds me of a story").
- local connectedness - contributions are fit to the preceding turn.
- progressivity - a topic is continued as long as it progresses (as long as there is new material, or until structural disfluencies occur).
- Interlocutors have a preference for gradual topic transition (Sacks, 1995), and sensitive topics can be broached in a stepwise and collaborative manner which displays the participants' mutual willingness to enter into it, and avoids dispreferred moves and other sudden shifts.

There are also constraints on the introduction of small talk within other types of talk. For example, in conversational frames in which there is an unequal power balance and some level of formality (e.g., job interviews), only the superior may introduce small talk in the medial phase of the encounter (Cheepen, 1988).

Other style constraints include the increased importance of politeness maxims and the decreased importance of Gricean "maximally informative communication" maxims, and the obligatory turn-taking mentioned above (one interlocutor cannot hold floor for the duration of the encounter).

In sum, as illustrated in Figure 4, relative to the strategies described above (and the relational dimensions they affect) small talk:

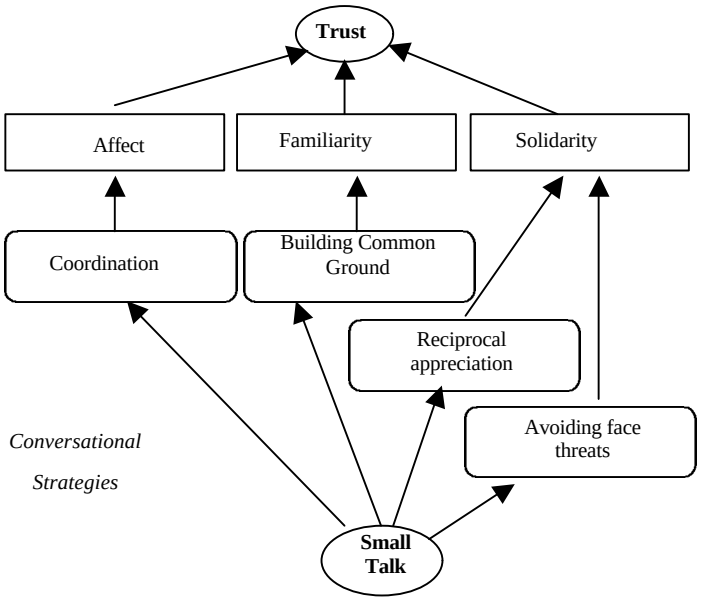

Figure 3: Small Talk Effects Strategies which Impact Dimensions of Relationship and Result in Trust

- Avoids *face threat* (and therefore maintains *solidarity*) by keeping conversation at a safe level of depth.
- Establishes *common ground* (and therefore increases *familiarity*) by discussing topics that are clearly in the context of utterance
- Increases *coordination* between the two participants by allowing them to sychronize short units of talk and nonverbal acknowledgement (and therefore leads to increased liking and positive *affect*).

4 Implementation of Social Language

Thus far we have described a model of the establishment of interpersonal relations in conversation, primarily based on the work of Svennevig, and augmented with insights from theories of politeness, phatic communication, non-verbal behaviors in conversation, and common ground. While we currently represent the model as a single entity per dyadic relationship, in fact people frequently form situation-specific representations of others and their relationships with them (they treat their colleagues differently in the gym than in the conference room). Thus, the above model may need to be replicated for different activity types, social roles, or contexts of use (Berscheid & Reis, 1998). The maintenance and integration of such a network of relational models is currently beyond the scope of this work, but provides an interesting area for future research. We next outlined 3 strategies for how interpersonal relationships may be established, maintained and changed through manipulating face work, common ground, and coordination. And we discussed one kind of social language, small talk, that can achieve all three strategies.

The theoretical structure that we have built is quite a complex one, and in this next section we turn to the implementation of the model in a real-time embodied conversational agent. We first briefly discuss the role of embodiment, and the particular embodiment that will be used, with a system architecture giving an overview of the agent. We then turn to the novel aspect of the current work, the user modeling and discourse planner sub-modules of the decision module, linking features of the implementation to the model of social language that we described above.

4.1 Embodiment in Conversational Agents

Since so many of the signs of interpersonal warmth are nonverbal, since face-to-face conversation is the medium of choice for the development of interpersonal relationships, and since nonverbal behaviors are essential to aspects of the management of dialogue, embodied conversational agents (ECAs) are particularly well-suited for use as the interfaces to social language systems. ECAs look like humans, can engage a user in real-time, multimodal dialogue, using speech, gesture, gaze, posture, intonation, and other nonverbal channels to emulate the experience of human face-to-face interaction (Cassell, Sullivan, Prevost, & Churchill, 2000). They rely on just-barely-conscious skills that humans cultivate from before the time they know how to speak: looking at one's conversational partner, giving feedback, using the timing among behaviors to convey meaning over and above the meaning of the words.

4.1.1 REA

REA is a real-time, multimodal, life-sized ECA, and her design is based on the FMBT model (Cassell, et al., 1999; Cassell, Bickmore, Vilhjálmsson, & Yan, 2000), pronounced *fembot*:

F. Division between Propositional and Interactional **F**unctions

Contributions to the conversation are divided into *propositional functions* and *interactional functions*. In short, the interactional discourse functions are responsible for creating and maintaining an open channel of communication between the participants, while propositional functions shape the actual content. Both functions may be fulfilled by the use of a number of available communication modalities. This feature is essential for the model of social relations as Rea can simultaneously pursue the prepositional goal of conveying new information, and the interactional goal of increasing familiarity.

M. Modality

Both verbal and nonverbal modalities are responsible for carrying out the interactional and propositional functions. It is not the case that the body behaviors are redundant. The use of several different modalities of communication - such as hand gestures, facial displays, eye gaze, and so forth - is what allows us to pursue multiple goals in parallel, some of a propositional nature and some of an

interactional nature. In addition, nonverbal behaviors are particularly important for coordination, and hence the feeling of "being in synch", and some particular nonverbal behaviors (such as feedback nods) act as markers of reciprocal appreciation, thereby increasing solidarity.

B. Behaviors are not functions

The same communicative function does not always map onto the same observed behavior. For instance, the interactional function of giving feedback could either be realized as a head nod or a short "mhm". The converse is also true - the same behavior does not always serve the same function. For example, a head nod could be feedback or equally well a salutation or emphasis on a word.

T. Time

Timing is a key property of human conversation, both within one person's conversational contributions, and between participants. Within one person's contribution, the meaning of a nod is determined by where it occurs in an utterance, to the 200 millisecond scale. The rapidity with which behaviors such as head nods achieve their goals emphasizes the range of time scales involved in conversation. While we have to be able to interpret full utterances to produce meaningful responses, we must also be sensitive to instantaneous feedback that may modify our interpretation and production as we go, and may allow users to coordinate their segments of talk with the agent.

Our FMBT conversational framework allows us to pursue multiple conversational goals in parallel (such as task goals and interpersonal goals), employ multiple modalities in parallel to achieve those goals (such as hand gestures and speech), and to use precise timings to achieve coordination among participants.

The REA embodied conversational agent 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 5). 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.

Figure 4: User interacting with Rea

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 (Stone & Doran, 1997), and extended to synthesize redundant and complementary conversational hand gestures, generates Rea's responses. Figure 6 shows

the architecture for the Rea system.

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 language normally occurs.

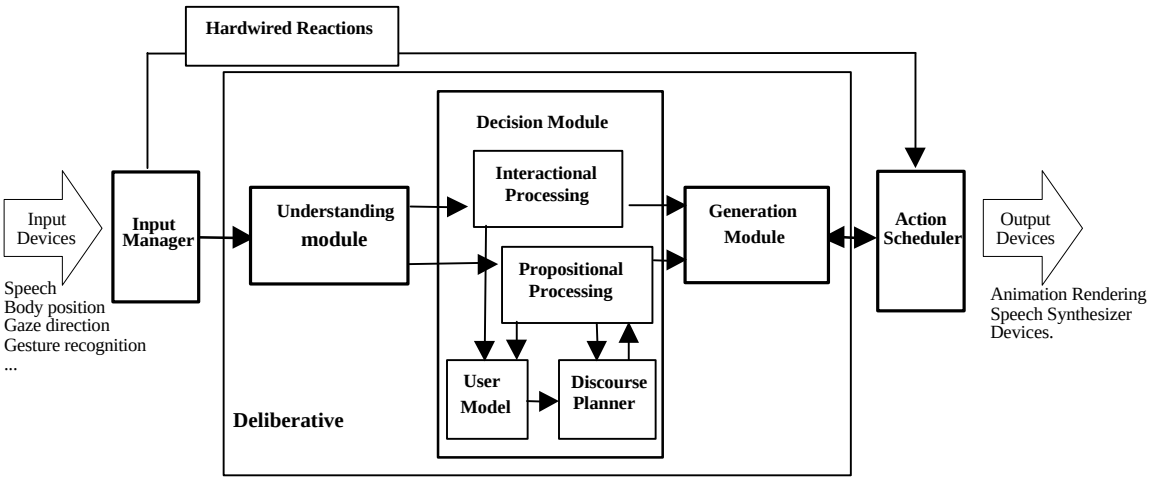

Figure 4: Rea System Architecture

4.2 User Modeling of Human-Computer Relationships

The social relationship between the user and a computer agent represents an underexplored aspect of user modeling. In our current implementation, we restrict ourselves to three of Svennevig's five relational dimensions--familiarity/depth, familiarity/breadth, and solidarity (Svennevig, 1999) -- each represented as a scalar ranging from zero to one, with increasing values representing increasing closeness. These elements of our user model are updated dynamically during the interaction with the user. We additionally have coordination or synchronization as an overall system goal, which contributes to the general positive dimension of liking. In fact, the current implementation does some amount of assessing user state, and adapting to it, but also engages in attempts to *change* user state – choosing behaviors that are intended ultimately to increase the user's trust in the system.

In the implementation, conversational topics are represented as objects which include measures of minimum and maximum 'social penetration' or invasiveness as two of their attributes. For example, talking about the weather does not represent a very invasive topic, whereas talking about finance does. Given these representations, depth of familiarity is updated based on the maximum invasiveness of all topics discussed, and breadth of familiarity is updated based on the number of topics discussed. The model of user solidarity with the system should ultimately be updated based on the similarity of the user's and agent's values and beliefs. However, for the moment we do not have access to these long-term features of the user. So, since solidarity has also been observed to increase based on the number of interactions two individuals have, our current model simply updates solidarity as a function of the number of turns of conversation that the user and agent have engaged in.

More formally, if $\boldsymbol{T} = \{t_1, t_2, \ldots t_J\}$ is the set of possible conversational topics the agent can discuss, $\boldsymbol{T_H}$ Í $\boldsymbol{T}$ is the set of topics already discussed, $\boldsymbol{T_C}$ Í $\boldsymbol{T}$ is the current set of topics under discussion, D_{MIN}:T®$0..1$ and D_{MAX}:T®$0..1$ represent the minimum and maximum social penetration for a topic (depth of familiarity), respectively, N_{moves} is the number of conversational moves made by the agent thus far in the

conversation and $\boldsymbol{M} = \{m_1, m_2, \ldots m_K\}$ is the set of possible conversational moves the agent can make, then the relational model is updated as follows.

$$FamiliarityDepth = \frac{Maximum(\{D_{MIN}(i) \mid i \in T_H \cup T_C\})}{Maximum(\{D_{MAX}(j) \mid j \in T\})}$$

$$solidarity = \frac{N_{moves}}{|M|}$$

$$FamiliarityBreadth = \frac{|T_H|}{|T|}$$

One final element of our user model is a set of topics $\boldsymbol{T_R} \acute{I} \boldsymbol{T}$ which are relevant to the user throughout the conversation. This set is initialized to topics regarding readily apparent features of the system and the immediate context of utterance -- the setting the user will be in when using the system -- such as REA's appearance and voice, the projection screen and microphone, the lab the system is situated in, MIT, Cambridge, Boston, the weather outside, etc. This set defines the topics that can readily be discussed with anyone who walks up to the system for the first time, and is thus important for determining topics for small talk, to increase common ground. Currently this set is not updated during use of the system, but ideally it would be expanded as more is learned about the user.

Following the model previously presented for face threat, the degree of threat for any given move m_i is computed as a scalar quantity based on the relational model as follows, given that $A:\boldsymbol{M} \mathbin{®} 2^{\boldsymbol{T}}$ is the set of topics a move is "about", $TC:< 2^{\boldsymbol{T}}, 2^{\boldsymbol{T}}> \mathbin{®} 0..1$ returns the degree of "topic coherence" between two sets of topics, ranging from 1 if the sets share any common members to 0 if the two sets of topics have nothing in common, $\boldsymbol{S} = \{'STORY', 'QUERY', 'STATEMENT'\}$ is the set of possible speech acts the agent can make, and SA: $\boldsymbol{M} \mathbin{®} \boldsymbol{S}$ provides the class of speech act for a conversational move.

face threat (m_i, *familiarity/depth, familiarity/breadth, solidarity*) =

$$SP_{threat} \text{ x } G_{SPthreat} + SAI_{threat} \text{ x } G_{SAIthreat} + SAC_{threat} \text{ x } G_{SACthreat}$$

Where,

SP_{threat} = Maximum($\{D_{MIN}(m_i)$ - *FamiliarityDepth*, 0$\}$)

SAI_{threat} = if *solidarity* >= S_{MIN} then 0 else

if $SA(m_i)$ = *'STORY'* then SA_{STORY}

else if $SA(m_i)$ = *'QUERY'* then SA_{QUERY}

else $SA_{STATEMENT}$

$SAC_{threat} = 1 - TC(A(m_i), \boldsymbol{T_C})$

$G_{SPthreat}$, $G_{SAIthreat}$, $G_{SACthreat}$ are constant gains

SA_{STORY}, SA_{QUERY} , $SA_{STATEMENT}$ are constants describing the degree of threat from telling a story, asking a question, or making a statement, respectively, if an appropriate level of solidarity (S_{MIN}) has not been established.

4.3 Discourse Planning for Mixed Task and Social Dialog

Conversation to achieve social goals (such as small talk) places many theoretically interesting demands on dialog systems, many of which have not been adequately – or at all – addressed by existent approaches to discourse planning. A discourse planner for social talk must be able to manage and pursue multiple conversational goals (Tracy & Coupland, 1991), some or all of which may be persistent or non-discrete. For example, in casual small talk, where there are apparently no task goals being pursued, interlocutors are conscious, nevertheless, of multiple goals related to conversation initiation, regulation and maintenance (Cegala, et al., 1988). Even in "task-oriented" interactions, speakers may also have several interpersonal goals they are pursuing, such as developing a relationship (e.g., befriending, earning trust) or establishing their reputations or expertise. It is not sufficient that a discourse planner work on one goal at a time, since a properly selected utterance can, for example, satisfy a task goal by providing information to the user while also advancing the interpersonal goals of the agent. In addition, many goals, such as intimacy or face goals, are better represented by a model in which degrees of satisfaction can be planned for, rather than the discrete all-or-nothing goals typically addressed in AI planners (Hanks, 1994). The discourse planner must also be very reactive, since the user's responses cannot be anticipated. The agent's goals and plans may be spontaneously achieved by the user (e.g., through volunteered information) or invalidated (e.g., by the user changing his/her mind) and the planner must be able to immediately accommodate these changes.

The action selection problem (deciding what an autonomous agent should do at any point in time) for conversational agents includes choosing among behaviors with an interactional function such as conversation initiation, turn-taking, interruption, feedback, etc., and behaviors with a propositional function such as conveying information. Within Computational Linguistics, the dominant approach to determining appropriate propositional behaviors has been to use a speech-act-based discourse planner to determine the semantic content to be conveyed. Once the content is determined, other processes (text generators) are typically used to map the semantic representations onto the words the agent actually speaks. This approach to discourse planning is based on the observation that utterances constitute speech acts (Searle, 1969), such as requesting, informing, wanting and suggesting. In addition, humans plan their actions to achieve various goals, and in the case of communicative actions, these goals include changes to the mental states of listeners. Thus, this approach uses classical "static world" planners (e.g., STRIPS (Fikes & Nilsson, 1971)) to determine a sequence of speech acts which will meet the agent's goals in a given context. One of the major advantages of plan-based theories of dialog is that language can be treated as a special case of other rational non-communicative behavior. Such an approach, however, cannot account for phenomena such as those described above in which many non-discrete factors need to be traded off in determining the next best action for a social agent to tak. Discrete reasoning yields a greatly underspecified solution for an agent which must reason about face threats, power, solidarity, and relative goal priorities, especially in social talk in which almost any topic can be raised at any given time, but at varying costs and benefits to the initiating agent.

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 the requirements to work towards the achievement of multiple, non-discrete goals in a dynamically changing environment, we have moved away from static world discourse planning, and are using an activation network-based approach based on Maes' *Do the Right Thing* architecture (Maes, 1989). In our implementation of this architecture, each node in the network represents a conversational move that REA can make, and links between the nodes represent various enabling and disabling conditions which hold among the moves (e.g., talking about the Boston real estate market introduces the topic of real estate thereby making it easier to introduce other real estate topics; see Figure 7).

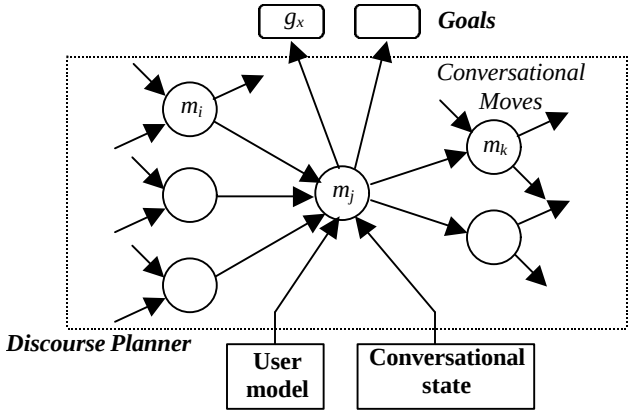

Figure 4: Conversational Moves in the Activation Network

Planning takes place as a spreading activation process that uses information from the current state of the conversation and relational model to determine which moves are more likely to succeed, along with information from the task goals so that REA can prioritize her possible moves to ensure that she addresses goals based on their relative importance, and adapted to her model of the user. Plans, ordered sequences of conversational moves, are represented as paths in the network, and are determined via the spreading activation mechanism. Although planning is performed before a move is selected, the mechanism does not return a complete plan like classical planners do. Instead, moves are selected for execution as necessary to achieve the unsatisfied task goals. During the activation process, energy is moved backward from the task goals to moves which directly lead to their achievement, from there to moves which enable those moves, and so on. Forward propagation is simultaneously performed by flowing energy into moves which can immediately be performed given the conversational state and relational model, and from there to moves which are enabled by those moves, and so on. The resulting paths in the network with the highest activation energy are thus those which are executable and best lead to satisfaction of the task goals.

Since the role of goals is to simply inject energy into moves which lead to their achievement, the network can straightforwardly be extended to work towards the achievement of non-discrete goals by simply varying the amount of energy the goals provide based not only on their relative priority, but on the difference between their current and desired degrees of satisfaction (the larger the discrepancy the more energy they provide). The pursuit of multiple goals can also be handled in parallel, with goal conflicts handled by inconsistent moves sending negative activation energy to each other. In addition, if a move does not produce an expected result, or if the conversational or relational states change in unanticipated ways, rather than re-planning (as a classical planner would do), the next best network path is automatically selected for execution. Thus, this architecture supports our requirements for planning to achieve multiple, non-discrete goals in a dynamic environment.

In addition, this architecture provides many features that are of particular use to designers of conversational agents. A discourse planner modeled in this manner provides enormous flexibility in designing agents whose conversational behavior vary by degree of goal-directedness, politeness, coherence, relevance or even deliberation (vs. opportunism), simply by changing the appropriate numeric gains controlling the amount of activation energy propagated under certain conditions. Since the spreading activation process incrementally estimates the best path to take through the network, it represents a form of "anytime" planner that can be stopped at any time to provide the best action to execute, although the longer it is allowed to run the better the result (Drummond & Bresina, 1990). Thus, the architecture provides the capability to transition smoothly from deliberative, planned behavior to

opportunistic, reactive behavior by varying the length of time the activation energy propagation algorithm runs.

4.3.1 Activation Network Architecture

We have adapted Maes' architecture for discourse planning in order to support mixed task and social dialog in REA, to adapt to the model of the user, and to dynamically attempt to change the user’s relationship with the system. 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 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:

- From Maes:

    1. 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.

    2. 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. Moves whose preconditions are satisfied by the user model and conversational state are given an increment of "forward chaining" energy. The move A which, when selected, will enable another move B, passes some of its activation energy forward from A to B. In addition, a move A which helps satisfy the preconditions of move B causes activation energy to flow from B to A, providing a "backward chaining" flow of energy.

- To deal with face threat avoidance:

    3. Face threat -- Moves which are expected to cause face threats to the user are dis-preferred, including face threats due to social penetration theory violations, speech act type or topic incoherence.

    4. Familiarity/Depth enablement -- In order to avoid face threats due to social penetration theory violations, REA can plan to perform small talk in order to "grease the tracks" for task talk, especially about sensitive topics like finance. To support this, energy is passed from moves whose familiarity/depth preconditions are not satisfied to those moves which would satisfy these preconditions if selected.

- To deal with topic management:

    5. Relevance -- Moves which involve topics in the list of topics known to be relevant to the user are preferred.

    6. Topic enablement -- 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). Thus, energy is propagated from

moves whose topics are not currently active to moves whose topics would cause them to become current.

More formally, given the set of agent goals $\boldsymbol{G} = \{g_1, g_2, \ldots\}$, the degree of satisfaction, $S_G:\boldsymbol{G}\rightarrow 0..1$, and priority, $P_G:\boldsymbol{G}\rightarrow 0..1$, for each goal, each move is assigned the following activation energy during each update cycle.

$$\boldsymbol{a}_i^0 = 0$$

$$\boldsymbol{a}_i^t = decay(\boldsymbol{a}_i^{t-1}) + E_{GOAL}(m_i) * G_{GOAL} + E_{ENABLED}(m_i) * G_{ENABLED} + E_{FORWARD}(m_i) * G_{FORWARD} + E_{BACKWARD}(m_i) * G_{BACKWARD} + E_{RELEVANCE}(m_i) * G_{RELEVANCE} + E_{FACETHREAT}(m_i) * G_{FACETHREAT} + E_{TOPICENABLE}(m_i) * G_{TOPICENABLE} + E_{SPENABLE}(m_i) * G_{SPENABLE}$$

Where,

G_{GOAL}, $G_{ENABLED}$, $G_{RELEVANCE}$, $G_{TOPICENABLE}$, $G_{SPENABLE,}$ $G_{FORWARD}$, and $G_{BACKWARD}$ are gain constants (≥ 0), and $G_{FACETHREAT}$ is a negative gain constant (≤ 0). Modification of these gains allows the agent to be made more or less goal-oriented (by changing G_{GOAL}), more or less polite (by changing $G_{FACETHREAT}$) or more or less deliberate in how topics are advanced (by changing $G_{TOPICENABLE}$).

$$decay(\boldsymbol{a}_i^t) = \frac{|M| \times p}{\sum_{j \in M} \boldsymbol{a}_j^t}$$

π is a normalization constant which controls the total amount of energy available in the network (the 'mean level of activation').

$$E_{GOAL}(m_i) = \sum_{g_j \in C_G(m_i)} (1 - S_G(g_j)) * P_G(g_j)$$

$C_G:M\rightarrow 2^G$ is the set of goals that a move directly contributes to the satisfaction of.

$E_{ENABLED}(m_i) = 1$ if all logical preconditions of the move are satisfied, 0 otherwise.

$$E_{FORWARD}(m_i) = \sum_{m_j \in M_{ENABLES}(m_i)} \boldsymbol{a}_j^{t-1}$$

$$E_{BACKWARD}(m_i) = \sum_{m_k \in M_{ENABLEDBY}(m_i)} \boldsymbol{a}_k^{t-1}$$

$M_{ENABLES}:M\rightarrow 2^M$ is the set of moves which have at least one logical precondition directly satisfied through the execution of a given move, and $M_{ENABLEDBY}:M\rightarrow 2^M$ is the inverse (the set of moves which, when executed, satisfy at least one logical precondition of the given move).

$$E_{RELEVANCE}(m_i) = TC(A(m_i), T_R)$$

$$E_{FACETHREAT}(m_i) = \text{facethreat}(m_i, \text{FamiliarityDepth}, \text{FamiliarityBreadth}, \text{solidarity})$$

$$E_{TOPICENABLE}(m_i) = \sum_{m_j \in M \mid A(m_j) - A(m_i) \neq \{\}} TC(A(m_i), A(m_j)) \times a_j^{t-1}$$

$$E_{SPENABLE}(mi) = \sum_{m_j \in SP_{ENABLES}(m_i)} a_j^{t-1}$$

$$SP_{ENABLE}(m_i) = \{m_j \mid m_j \in M \wedge D_{MOVEMIN}(m_j) \geq \text{FamiliarityDepth} \wedge D_{MOVEMIN}(m_j) \leq D_{MOVEMAX}(m_i) \wedge \text{FamiliarityDepth} < D_{MOVEMIN}(m_j)\}$$

$$D_{MOVEMAX}(m_i) = \text{Maximum}(\{D_{MAX}(x) \mid x \in A(m_i)\})$$
$$D_{MOVEMIN}(m_i) = \text{Minimum}(\{D_{MIN}(x) \mid x \in A(m_i)\})$$

This last factor propagates energy from a move which is currently dis-preferred because of a social penetration theory violation to moves which could enable it by increasing *FamiliarityDepth* when executed.

Given the above activation energy update rule, a threshold of activation, θ, and a threshold decrement, $0 < q_{DECREMENT} < 1$, planning in the network proceeds as follows.

$q \neg q_{INITIAL}$

while a move has not been selected do

compute a_i *for all moves*

select move m_i *with maximum* a_i *such that* $a_i > q$ *and* $E_{ENABLED}(m_i) = 1$

if no such move is found then $q \neg q * q_{DECREMENT}$

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. This strategy may seem to indicate the opposite of user modeling but, in practice, much human-human small talk proceeds along similar lines and as described above, the tight temporal coordination of units is actually more important than content.

At each step in the conversation in which Rea has the floor (as tracked by a conversational state machine in Rea's Reaction Module (Cassell, Bickmore, Vilhjálmsson, & Yan, 2000)), the discourse planner is consulted for the next conversational move to initiate. At this point, activation values are incrementally propagated through the network (following the algorithm above) until a move is selected whose preconditions are satisfied and whose activation value is above the specified threshold. Moves are

executed differently depending on their type. Task queries consist of REA question/user replay pairs; task and small talk statements consist of a REA statement turn only; and small talk stories and queries consist of a REA contribution/optional user response/REA idle response triples.

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 (Clark, 1996).

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).

Note that this implementation is a simplification of Maes' architecture in that it currently assumes information in the conversational state is monotonic, thus goal protection and action conflicts are not currently dealt with. We also assume that each conversational move can only be used once in a given interaction and thus disable moves that have been executed by effectively removing them from the network. Finally, given that the threshold of activation, q, is decreased on each update cycle, $q_{DECREMENT}$ controls the amount of deliberation the network performs by controlling the number of update cycles executed before a move is selected. As long as $q_{DECREMENT} < 1$ the algorithm will eventually yield a result unless there are no moves available whose logical preconditions are satisfied. In practice, a $q_{DECREMENT}$ of 0.1 (as used by Maes) along with $q_{INITIAL}=3p$ *and* $p=1/|M|$ results in move selection after just a few update cycles.

4.3.2 Related Work in Activation Network-Based Planning

Goetz recast Maes' networks as connectionist Hopfield networks which perform pattern recognition. In the process he discovered several interesting constraints and shortcomings in Maes' networks, but most importantly demonstrated that if certain non-linearities are added to the update rules the behavior of the network became more stable with respect to persistently pursuing a given plan (Goetz, 1997).

A more recent, probabilistic, reformulation of this approach to planning was taken by Bagchi, et al. (Bagchi, Biswas, & Kawamura, 1996), in which the network consists of actions and explicitly represented propositions which are pre- and post-conditions of the actions. In this architecture, the activation values associated with propositions reflect the probability of their being true, while the values associated with actions reflect their expected utility. The process of spreading activation is used to incrementally estimate these probabilities and utilities using calculations local to each node in the network. In this approach, the action with the highest utility is selected at the end of each update cycle for execution. We have not adopted this probabilistic approach given the extreme subjectivity involved in estimating the various costs and probabilities which comprise the network, and since it has not been extended to deal with non-discrete goals or propositions yet. However, we find it a promising direction for future work.

5 Example Interactions

We feel that the original goals of developing a discourse planner capable of working towards multiple, non-discrete goals in a dynamic environment have been satisfied by the model and implementation presented, and that it meets the needs of discourse planning for mixed task and social dialog to assess and adapt to user relational state.

In our real estate domain we have several task goals--such as finding out information about the user's desired location, price range, house size, and amenities--with varying priorities (price and location are most important). The interaction of these goals with the dynamically changing user model yields what we believe to be fairly natural conversational behavior for this domain. With minimal tuning of the network

gains Rea can be made very goal-oriented or very chatty, although finding desired in-between behaviors can require some tuning effort. We have found that as long as $G_{SACthreat}$ is kept high (maintaining coherence) and $G_{RELEVANCE}$ is kept above zero (maintaining some user relevance) the resulting conversational behavior is natural and believable.

There are some limitations of this approach with respect to other forms of planning, however. In the current model the moves in the network represent 'ground level' actions rather than abstract schemata, limiting the flexibility and scalability of the approach relative to classical hierarchical planners (something we plan to address in future work). There are also no guarantees of correctness or completeness of the plans produced; the spreading activation approach is a heuristic one. Finally, it is unclear how activation network based planners could deal with the very complex goal interactions or temporal constraints that many classical planners have been designed to handle.

In what follows we reproduce some actual output from Rea in conversation with a user (user responses are only shown in positions in which they affect the selection of subsequent joint projects). The first example illustrates Rea engaging in baseline small talk.

	Move	Fam/D	Fam/B	Solidarity
1.	How about this weather?	0.00	0.00	0.00
2.	I think winters in Boston are awful.			
3.	How do you like Boston?			
4.	I have lived in Boston all my life. Come to think of it, I have lived inside this room all of my life. It is so depressing.			
5.	Boston is certainly more expensive than it used to be.	0.50	0.19	0.17
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?	0.60	0.29	0.30
10.	You know, I keep showing the researchers here the same houses, over and over again. Maybe one day I will get lucky.			
11.	Have you been in the Media Lab before?			
12.	Do you know that the Media Lab is going to expand into another building. Things are really going well for the researchers here.			
13.	It is pretty cool do you think?			
14.	They are doing some crazy things in here.			
15.	I have shown houses to lots of students and faculty from M I T. But I always enjoy talking to them.	0.70	0.38	0.50
16.	Anyway, What can you afford?			
17.	What kind of down payment can you make?			
18.	Let me see what I have available.	0.90	0.43	0.57

Example 1. "Social Language REA"

In this example, REA opens with small talk moves regarding things she knows she can discuss with the user (1-5). After a few turns, enough familiarity has been established (simply by doing small talk) that

REA can move into task talk (6-9), aided by the transition move in turn 5 (linking Boston to real estate). However, before bringing up the topic of finance -- a topic that is potentially very face threatening for the user -- REA decides that additional familiarity needs to be established, and moves back into small talk (10-15). The move at turn 15 transitions the topic from the Media Lab back to real estate, enabling REA to then bring up the issue of how much the user is able to afford (16-17).

	Move	Fam/D	Fam/B	Solidarity
1.	So, where would you like to live?	0.00	0.00	0.00
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.	0.50	0.19	0.20

Example 2. "Task REA"

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

As a final example, the importance of task goal satisfaction can be reduced (relative to the baseline case) yielding the following dialogue in which Rea rambles on at length before getting down to business.

	Move	Fam/D	Fam/B	Solidarity
1.	How about this weather?	0.00	0.00	0.00
2.	I think winters in Boston are awful.			
3.	How do you like Boston?			
4.	Boston is certainly more expensive than it used to be.	0.40	0.14	0.13
5.	So, where would you like to live?			
6.	How many bedrooms do you need?	0.50	0.24	0.20
7.	You know, I keep showing the researchers here the same houses, over and over again. Maybe one day I will get lucky.			
8.	Have you been in the Media Lab before?			
9.	Do you know that the Media Lab is going to expand into another building. Things are really going well for the researchers here.			
10.	It is pretty cool do you think?			
11.	They are doing some crazy things in here.			
12.	I have shown houses to lots of students and faculty from MIT. But I always enjoy talking to them.			
13.	I really love it here at MIT. Do you?			
14.	I suppose you have heard about all of the hacks at MIT? My favorite was when they dressed the dome up to look like a robot.			
15.	It is pretty intense, do you think?			
16.	Are you tired of demos yet?			
17.	That microphone is terrible. I hate using			

	those things.			
18.	Sorry about my voice. This is some engineers idea of natural sounding.	0.60	0.43	0.60
19.	Anyway, what can you afford?			
20.	What kind of down payment can you make?			
21.	Do you need access to the subway?			
22.	Is one bath enough?			
23.	Let me see what I have available.	0.90	0.48	0.73

Example 3. "Chatty REA"

When task goals become less important, topic coherence and other source of activation energy become the primary factors in deciding what Rea says next.

6 Evaluation

Thus far we have presented a rather weighty theoretical model to relate social language to interpersonal relations. And we have describing a rather ponderous formalism and implementation for generating social language to achieve interpersonal goals in embodied conversational agents. Does the implementation float under all of this weight? In the previous section we discussed whether the implementation stands up, and its current limitations. Here we address whether small talk produced by an ECA in a sales encounter has any effect whatsoever on computer-human interaction.

In order to evaluate whether an ECA's social language can actually build trust, solidarity, and interpersonal closeness 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. Rea is entirely implemented. However, for the purpose of the experiment, Rea was controlled by a human wizard, following scripts identical to the output of the planner -- but not dependent on network traffic, automatic speech recognition or computational vision (Oviatt, 1996). The study was a between subjects design with subjects randomly assigned either to a version of REA which used only task-oriented dialogue (TASK condition) or to an identical version which also included the social dialogue (SMALLTALK condition).

The questions we asked concerned the effects of modeling user trust, user interpersonal distance and user comfort with an interaction, and using social language to manipulate those dimensions in users during the interaction. Remember that our implementation of our model gauges the threat of particular topics and uses social talk to increase user comfort before introducing them; the implementation explicitly tries to raise trust (increasing solidarity, familiarity, and liking) by building common ground, minimizing face threat, coordinating with the user, acknowledging the user's contributions.

Our hypotheses for this empirical evaluation follow from the literature on small talk and on interpersonal relations among humans. Because trust is an outcome from the strategies that we intended Rea to implement with her small talk, we expected subjects in the SOCIAL condition to trust Rea more. We also expected them to feel closer to Rea, like her more, and feel that they understand her and were understood by her more than in the TASK condition. We 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.

6.1 Experimental Method

Subjects. 31 people participated in the experiment (58% male and 42% female). Subjects were primarily students, were recruited through ads on several college campuses, and were compensated for their participation.

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. All three manipulation check variables were highly significant. For example, there was a significant difference ($F = 11.2$; $p < .002$) such that users believed that Rea got down to business more quickly in the task-only condition than in the small talk condition.

6.2 Measures

Trust was measured by a standardized trust instrument (Wheeless & Grotz, 1977). The measurement was calculated by asking subjects to rate Rea on a number of Likert scales where they had to place her between, for example, candid and deceptive, benevolent and exploitative, and so forth (alpha = .88 (Nass & Lee, 2000)).

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 included on the questionnaire that subjects responded to at the end of the experiment a standard set of questions that are commonly used to judge extrovertedness and introvertedness (Nass & Lee, 2000).

Extrovertedness was an index composed of seven Wiggins (Wiggins, 1979) extrovert adjective items: Cheerful, Enthusiastic, Extroverted, Jovial, Outgoing, and Perky. It was used for assessment of the subject (Cronbach's alpha = .94).

Introvertedness was an index composed of seven Wiggins (Wiggins, 1979) introvert adjective items: Bashful, introverted, Inward, Shy, Undemonstrative, Unrevealing, and Unsparkling. It was used for assessment of the subject (alpha = .83).

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 *initiaters* (above the mean on number of user-initiated utterances). To our surprise, this measure turned out to be independent of intro/extroversion, and to not be predicted by these latter variables (Pearson r = 0.053). We hypothesized that those subjects who were interrupted by Rea would be less happy with her performance, since she would not let them finish their utterances.

7 Results

Full factorial single measure ANOVAs were run, with CONDITION and PERSONALITY as independent variables. The most striking results obtained were main effects for Rea's perceived knowledgeability, and informedness – in both cases, the small talk condition scored significantly higher on these dimensions – and interactions between intro/extroversion and trust (and intro/extroversion and a number of other positive variables), and interactions between initiative/passivity and engagement (and a number of other positive variables).

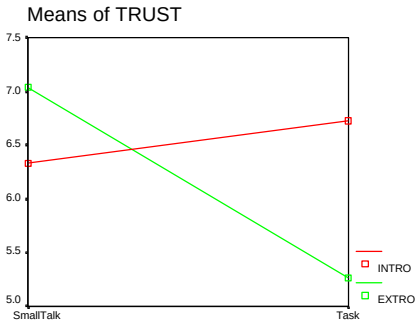

Figure 8: Trust Estimation by introverts & extroverts

Figure 8 shows the interaction between intro/extroversion and trust ($F=5.0$; $p<.05$). These results indicate that small talk had essentially no effect on the trust assessment of introverts. However, this kind of social talk 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. One extrovert in the SmallTalk condition commented

> I thought she was pretty good. You know, I can small talk with somebody for a long time. It's how I get comfortable with someone, and how I get to trust them, and understand how trustworthy they are, so I use that as a tool for myself.

An extrovert in the TaskOnly condition, on the other hand, remarked

> Great for people on the go and are looking for an easy way to find information about apartments. You lose some of that personal interaction with her. Yes, she was knowledgeable about what was out there. Yes, she asked the right questions about what I wanted, but in terms of getting to know her, that didn't take place. I felt like I was talking to a machine vs. talking to a person. That was the only thing that kind of threw me off. I would have liked her to ask more questions about what I like, getting to know more who I am, that would have made me more comfortable, at least in this scenario.

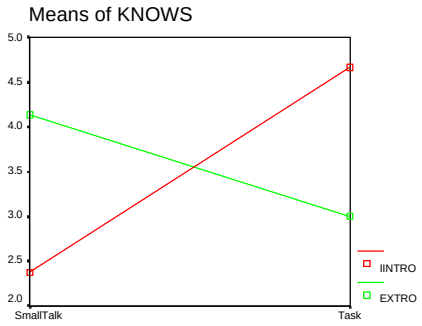

Figure 9: How well REA knew introvert & extrovert users

Extroverts said they felt that REA knew them and their needs better in the SmallTalk condition, while introverts said that REA knew them better in the Task condition ($F=4.4$; $p<0.05$) (see Figure 9). Extroverts also said they felt that they knew REA better in the SmallTalk condition, while introverts said that they knew REA better in the Task condition ($F=5.3$; $p<0.03$) (see Figure 10).

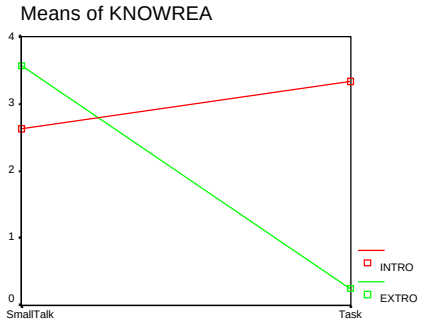

Figure 10: How well introvert & extrovert users knew REA

Extroverts also felt the interaction was more natural (F=4.0; p<0.06), satisfying (F=9.6; p<0.005) (see Figure 11) and successful (F=5.4; p<0.03) with SmallTalk, while introverts said the same of the Task condition. On the other hand, testifying to the utility of such a system even for introverts, during the debrief session, when asked about the naturalness of the interaction, one introvert user in the SmallTalk condition commented "It was really well done. I was thinking that if she can do it, then any person can learn how to chit chat."

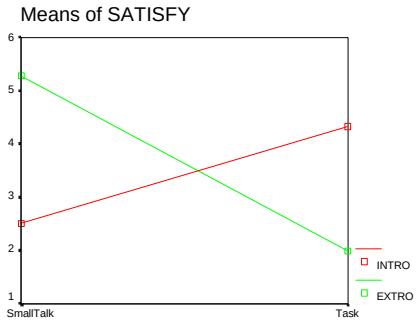

Figure 11: How satisfying the interaction was by introverts & extroverts

Finally, extroverts said that REA was more credible in the SmallTalk condition, while introverts felt she was more credible in the Task condition (F=3.4; p<0.08).

As noted above, to our surprise, initiative taking was not correlated with intro-/extroversion. Full factorial ANOVAs were therefore again performed on all measures, with CONDITION and INITIATIVE as dependent variables. Figure 12 shows the interaction between initiative/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=3.9; p<0.06).

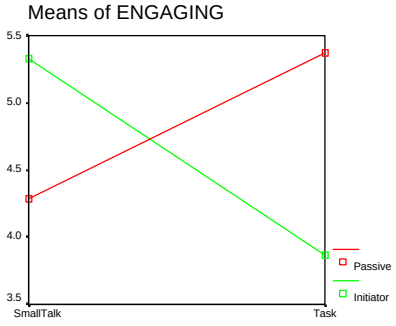

Figure 12: Engagement by initiaters vs. passive speakers

Likewise, more active users felt as if the interaction were more interesting (F=5.2; p<0.05), as if Rea came to know them better (F=4.4; p<0.05), that they knew Rea better (F=14.3; p<0.001) (see Figure 13), and that Rea was more of an expert (F=3.5;p<0.08) when she used small talk.

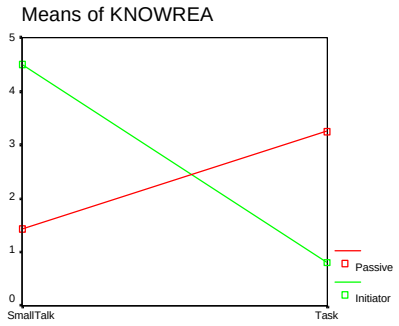

Figure 13: How well initiating vs. passive users felt they knew REA

These results concerning initiative-taking vs. passive speakers were surprising. Remember that the initiators were the subjects who most often experienced being interrupted by Rea, which led us to hypothesize that initiators would be less satisfied with the interaction. In all of these cases, however, 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. One thought, for example, is that trust is not implicated in the price of the apartment, but in the price the *realtor* demands. In order to examine this issue, we need to more directly target the realtor's involvement in the price. For example, we might ask "do you think the realtor asked you for a fair price for this apartment."

8 Discussion and Future Work

In this article we have examined a new aspect of user modeling – how to assess, adapt to, and potentially change, the "interpersonal" relationship that a user feels with a system. We set out to look at a fairly short-term user feature within this domain – the level of interpersonal closeness that the user feels – but we discovered in our evaluation of the system that there is an interaction between this short term feature and two longer-term features – the user's personality (introverted vs. extroverted) and the user's dialogue style (initiating vs. passive).

The results of our evaluation underline the fact that many people simply prefer a conversational partner who tries to get to know them, and who takes into account the interpersonal dimensions of the interaction. However, it is also clear that there are significant differences in reactions to the interaction depending on user disposition. This difference is exemplified by the following comment:

> REA exemplifies some things that some people, for example my wife, would have sat down and chatted with her a lot more than I would have. Her conversational style seemed to me to be more applicable to women, frankly, than to me. I come in and I shop and I get the hell out. She seemed to want to start a basis for understanding each other, and I would glean that in terms of our business interaction as compared to chit chat. I will form a sense of her character as we go over our business as compared to our personal life. Whereas my wife would want to know about her life and her dog, whereas I really couldn't give a damn.

Of course, as this comment also illustrates, one issue that must be addressed as well in extensions of this experimental paradigm is the sheer length of the dialogues in the social talk vs. task talk conditions. As it stands, social talk adds additional conversational turns and therefore time to the interaction. In the experimental protocol, including greeting, apartment showing and farewell (which were identical for both

conditions), small talk subjects engaged in 48 conversational turns while task talk subjects engaged in 36. The difference is not enormous, however one might think that solidarity would deepen simply because of time spent together. This factor might even explain why introverts are less comfortable with small talk, if they are less comfortable with talk in general. However, such a potential confound does not seem to explain results concerning dialogue style (initiating vs. passive), nor some of the particular interactions between personality and social talk. For example it is difficult to imagine length having an effect on the interaction between personality and credibility. More generally, it is a challenge to construct a task dialogue that is as long as a small talk one, without adding depth. Nevertheless this issue should be addressed in the future.

Remember that in the evaluation reported here judgments of introversion and extroversion were done on the basis of post-experiment questionnaires. And, while the system currently keeps an updated model of the user's interpersonal distance (familiarity/depth), the level of solidarity, and the range of topics shared between the system and user (familiarity/breadth), it does not model or adapt to the user's personality or discourse style. The interactions that we found between these dispositions and short-term user state, however, indicate that we might wish to model introversion and extroversion, dialogue initiative and dialogue passivity in such a way that these characteristics determine the direction of the interaction. The responses to 4 or 5 subtle questions could let us know whether the current user is the kind of person who will appreciate small talk or abhor it. Additional strategies for dynamically assessing the current state of the relationship with the user might also be developed, in addition to strategies for assessing the user's personality type (introvert/extrovert, active/passive, etc.), since these will affect the relational strategies that can successfully be used by an agent.

There are many other directions we are also considering for the future. For example, the current implementation only models the solidarity and familiarity dimensions of the computer-user relationship; the additional dimensions of affect and power have yet to be addressed dynamically during the interaction. In addition, we have so far only modeled the relational strategy of small talk, there are a large number of additional strategies that can be explored including ingratiation, explicit self-disclosure, humor, in-group talk, etc.

We have only begun to investigate the role of the body and nonverbal behavior in signaling trust and other relational strategies. Some literature suggests that the performance of the body differs in trusting vs. nontrusting states – for example, there is evidence that people are disfluent when they don't trust their conversational partner. If indeed users behave less effectively when the interpersonal dimension of the interaction has not been addressed, this is an additional incentive to model strategies for achieving interpersonal equilibrium (and additional incentive to use language as a dependent variable – to examine the role of lack of trust on the very nature of the interaction). We also expect the role of affect recognition and display to play a significant role in relational strategies, for example in order to show caring and empathetic behavior an agent must be attuned to the affective state of the user (Picard, 1997).

We have chosen to focus on the non-discrete aspects of dialogue planning, but task decomposition and discourse modeling, for example as performed in COLLAGEN (Rich & Sidner, 1997), must be brought back in and integrated in order for a relational agent to participate in non-trivial collaborative tasks with the user.

Although the model provides mechanisms for describing *how* an ECA may change its relationship with a user, it does not say anything about *why* it may want to do this. A more comprehensive theory may take yet another step back in the causal chain and determine the situations in which it is advantageous or disadvantageous to employ one of the relational strategies described here. An ECA may also have a wider range of social goals than to change its relationship with the user, and these will have to be integrated into the model of a fully socially intelligent agent, possibly along the lines of Pautler's model of social perlocutions (Pautler, 1998), or Castelfranchi and de Rosis' model of sincere assertions (Castelfranchi & de Rosis, 1999).

Finally, issues of privacy and ethics, while orthogonal to this work, should be investigated with respect to the development and deployment of relational agents. Users should know what cues a relational agent is using so that they can employ the same techniques for hiding their personality, goals, etc., that they would use with other people in similar circumstances, should they not want this information acquired by an agent. In addition, ethical questions such as when an agent should try to manipulate its relationship with the user, and what techniques it should be allowed to use, need to be answered before deployment of these technologies becomes widespread.

Social intelligence includes knowledge of when and how to use language to achieve social goals. As embodied conversational agents become ubiquitous, the ability for them to establish and maintain social relationships with us will become increasingly important. The study of how to constitute relationships through language will inform our growing ability to emulate aspects of humans in the service of efficient and effective interaction between humans and machines. In the meantime, we have demonstrated that it is possible to model dimensions of social relationships and realize change along those dimensions by using social language to accomplish interpersonal goals.

9 Acknowledgements

Research leading to the preparation of this article was supported by the National Science Foundation (award IIS-9618939), AT&T, and the other generous sponsors of the MIT Media Lab. Thanks to the current Rea team – Lee Campbell, Yukiko Nakano, Ian Gouldstone, Hannes Vilhjalmsson – for their development efforts, and to Diane Garros, realtor extraordinaire. Sincere thanks also to Dan Ariely, Cliff Nass, Candy Sidner, Matthew Stone and four anonymous reviewers for generous and helpful comments that improved the paper.

10 References

Andersen, P., & Guerrero, L. (1998). The Bright Side of Relational Communication: Interpersonal Warmth as a Social Emotion. In P. Andersen & L. Guerrero (Eds.), *Handbook of Communication and Emotion* (pp. 303-329). New York: Academic Press.

Andre, E., Muller, J., & Rist, T. (1996). The PPP Persona: A Multipurpose Animated Presentation Agent, *Advanced Visual Interfaces* pp. 245-247: ACM Press.

Ardissono, L., Boella, G., & Lesmo, L. (1999). Politeness and speech acts, *Proceedings of the Workshop on Attitudes, Personality and Emotions in User-Adapted Interaction at the 7th International Conference on User Modeling (UM '99)*, Banff, Canada.

Argyle, M. (1990). The biological basis of rapport. *Psychological Inquiry, 1*, 297-300.

Bagchi, S., Biswas, G., & Kawamura, K. (1996). Interactive task planning under uncertainty and goal changes. *Robotics and Autonomous Systems, 18*, 157-167.

Ball, G., & Breese, J. (2000). Emotion and Personality in a Conversational Agent. In J. Cassell, J. Sullivan, S. Prevost, & E. Churchill (Eds.), *Embodied Conversational Agents*. Cambridge, MA: MIT Press.

Berscheid, E., & Reis, H. (1998). Attraction and Close Relationships. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *The Handbook of Social Psychology* (pp. 193-281). New York: McGraw-Hill.

Beskow, J., & McGlashan, S. (1997). Olga: a conversational agent with gestures, *IJCAI 97*, Nagayo, Japan: Morgan-Kaufmann Publishers.

Brown, J.R., & Rogers, E.L. (1991). Openness, uncertainty and intimacy: An epistemological reformulation. In N. Coupland, H. Giles, & J.M. Wiemann (Eds.), *Miscommunication and problematic talk* (pp. 146-165). Newbury Park, CA: Sage.

Brown, P., & Levinson, S. (1978). Universals in language usage: Politeness phenomena. In E. Goody (Ed.), *Questions and Politeness: Strategies in Social Interaction* (pp. 56-311). Cambridge: Cambridge University Press.

Brown, R., & Gilman, A. (1972). The pronouns of power and solidarity. In P. Giglioli (Ed.), *Language and Social Context* (pp. 252-282.). Harmondsworth: Penguin.

Brown, R., & Gilman, A. (1989). Politeness theory and Shakespeare's four major tragedies. *Language in Society, 18*, 159-212.

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.

Cassell, J., & Bickmore, T. (2000). External Manifestations of Trustworthiness in the Interface. *Communications of the ACM, 43*(12), 50-56.

Cassell, J., Bickmore, T., Billinghurst, M., Campbell, L., Chang, K., Vilhjalmsson, H., & Yan, H. (1999). Embodiment in Conversational Interfaces: Rea, *CHI 99* pp. 520-527, Pittsburgh, PA.

Cassell, J., Bickmore, T., Vilhjálmsson, H., & Yan, H. (2000). More Than Just a Pretty Face: Affordances of Embodiment, *IUI 2000* pp. 52-59, New Orleans, Louisiana.

Cassell, J., Sullivan, J., Prevost, S., & Churchill, E. (2000). *Embodied Conversational Agents*. Cambridge: MIT Press.

Castelfranchi, C., & de Rosis, F. (1999). Which User Model do we need, to relax the hypothesis of 'Sincere Assertion' in HCI?, *Proceedings of the Workshop on Attitudes, Personality and Emotions in User-Adapted Interaction at the 7th International Conference on User Modeling (UM '99)*, Banff, Canada.

Cegala, D., Waldro, V., Ludlum, J., McCabe, B., Yost, S., & Teboul, B. (1988). A study of interactants' thoughts and feelings during conversation., *Ninth Annual Conference on Discourse Analysis*, Philadelphia, PA.

Cheepen, C. (1988). *The Predictability of Informal Conversation*. New York: Pinter.

Clark, H.H. (1996). *Using Language*. Cambridge: Cambridge University Press.

Dehn, D.M., & van Mulken, S. (in press). The Impact of Animated Interface Agents: A Review of Empirical Research. *International Journal of Human-Computer Studies, 51*.

Depaulo, B., & Friedman, H. (1998). Nonverbal Communication. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *The Handbook of Social Psychology* (pp. 3-40). Boston: McGraw-Hill.

Drummond, M., & Bresina, J. (1990). Anytime synthetic projection: Maximizing the probability of goal satisfaction, *AAAI-90* pp. 138-144.

Fikes, R., & Nilsson, N. (1971). STRIPS: A new approach to the application of theorem proving to problem solving. *Artificial Intelligence, 5*(2), 189-208.

Fogg, B.J. (1999). Persuasive Technologies, *Communications of the ACM, Vol. 42* (pp. 27-29).

Fogg, B.J., & Tseng, H. (1999). The Elements of Computer Credibility, *CHI 99* pp. 80-87, Pittsburgh, PA: ACM Press.

Goetz, P. (1997). *Attractors in Recurrent Behavior Networks*. PhD Thesis, State University of New York at Buffalo.

Goffman, I. (1967). On face-work, *Interaction Ritual: Essays on Face-to-Face Behavior* (pp. 5-46). New York: Pantheon.

Grice, P. (1989). *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Hanks, S. (1994). Discourse Planning: Technical Challenges for the Planning Community., *AAAI Workshop on Planning for Inter-Agent Communication*.

Jaworski, A., & Coupland, N. (1999). *The Discourse Reader*. London: Routledge.

Jefferson, G. (1978). Sequential aspects of storytelling in conversation. In J. Schenkein (Ed.), *Studies in the organization of conversational interaction* (pp. 219-248). New York: Academic Press.

Lester, J., Voerman, J., Towns, S., & Callaway, C. (1999). "Deictic Believability: Coordinating Gesture, Locomotion, and Speech in Lifelike Pedagogical Agents". *Applied Artificial Intelligence, 13*(4-5), 383-414.

Maes, P. (1989). How to do the right thing. *Connection Science Journal, 1*(3), 291-323.

Malinowski, B. (1923). The problem of meaning in primitive languages. In C.K. Ogden & I.A. Richards (Eds.), *The Meaning of Meaning* (pp. 296-336): Routledge & Kegan Paul.

Mark, G., & Becker, B. (1999). Designing believable interaction by applying social conventions. *Applied Artificial Intelligence, 13*, 297-320.

Moon, Y. (1998). Intimate self-disclosure exchanges: Using computers to build reciprocal relationships with consumers. Cambridge, MA: Harvard Business School.

Morkes, J., Kernal, H., & Nass, C. (1998). Humor in Task-Oriented Computer-Mediated Communication and Human-Computer Interaction, *CHI 98* pp. 215-216, Los Angeles, CA: ACM Press.

Nass, C., & Lee, K. (2000). Does Computer-Generated Speech Manifest Personality? An Experimental Test of Similarity-Attraction, *CHI 2000* pp. 329-336, The Hague, Amsterdam: ACM Press.

Oviatt, S. (1996). User-Centered Modeling for Spoken Language and Multimodal Interfaces. *IEEE MultiMedia, 1996*, 26-35.

Pautler, D. (1998). A Computational Model of Social Perlocutions, *COLING/ACL*, Montreal.

Picard, R. (1997). *Affective Computing*. Cambridge, MA: MIT Press.

Reeves, B., & Nass, C. (1996). *The Media Equation: how people treat computers, televisions and new media like real people and places*. Cambridge: Cambridge University Press.

Resnick, P.V., & Lammers, H.B. (1985). The Influence of Self-esteem on Cognitive Responses to Machine-Like Versus Human-Like Computer Feedback. *The Journal of Social Psychology, 125*(6), 761--769.

Rich, C., & Sidner, C.L. (1997). COLLAGEN: When Agents Collaborate with People, *Autonomous Agents 97* pp. 284-291, Marina Del Rey, CA.

Rich, E. (1979). User Modeling via Stereotypes. *Cognitive Science, 3*, 329-354.

Rickel, J., & Johnson, W.L. (1998). Animated Agents for Procedural Traning in Virtual Reality: Perception, Cognition and Motor Control. *Applied Artificial Intelligence, 13*(4-5), 343-382.

Rickenberg, R., & Reeves, B. (2000). The Effects of Animated Characters on Anxiety, Task Performance, and Evaluations of User Interfaces, *CHI 2000* pp. 49-56, The Hague, Amsterdam.

Sacks, H. (1995). *Lectures on Conversation*. Oxford: Blackwell.

Schneider, K.P. (1987). Topic selection in phatic communication. *Multilingua, 6*(3), 247-256.

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

Searle, J. (1969). *Speech Acts: An essay in the philosophy of language*: Cambridge University Press.

Spencer-Oatey, H. (1996). Reconsidering power and distance. *Journal of Pragmatics, 26*, 1-24.

Stone, M., & Doran, C. (1997). Sentence Planning as Description Using Tree-Adjoining Grammar, *ACL* pp. 198--205, Madrid, Spain: MIT Press.

Svennevig, J. (1999). *Getting Acquainted in Conversation*. Philadephia: John Benjamins.

Thorisson, K.R. (1997). Gandalf: An Embodied Humanoid Capable of Real-Time Multimodal Dialogue with People, *Autonomous Agents '97*.

Tracy, K., & Coupland, N. (1991). Multiple goals in discourse: An overview of issues. In K. Tracy & N. Coupland (Eds.), *Multiple goals in discourse* (pp. 1-13). Clevedon: Multilingual Matters.

Ward, N. (1997). Responsiveness in Dialog and Priorities for Language Research. *Systems and Cybernetics, Special Issue on Embodied Artificial Intelligence, 28*, 521--533.

Wheeless, L., & Grotz, J. (1977). The Measurement of Trust and Its Relationship to Self-Disclosure. *Human Communication Research, 3*(3), 250-257.

Wiggins, J. (1979). A psychological taxonomy of trait-descriptive terms. *Journal of Personality and Social Psychology, 37*(3), 395-412.