Chapter 1

SOCIAL DIALOGUE WITH EMBODIED CONVERSATIONAL AGENTS

Timothy Bickmore
Northeastern University, USA
bickmore@ccs.neu.edu

Justine Cassell
Northwestern University, USA
justine@northwestern.edu

Abstract The functions of social dialogue between people in the context of performing a task is discussed, as well as approaches to modelling such dialogue in embodied conversational agents. A study of an agent's use of social dialogue is presented, comparing embodied interactions with similar interactions conducted over the phone, assessing the impact these media have on a wide range of behavioural, task and subjective measures. Results indicate that subjects' perceptions of the agent are sensitive to both interaction style (social vs. task-only dialogue) and medium.

Keywords: Embodied conversational agent, social dialogue, trust.

1. Introduction

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 (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 neighbourhoods with good schools. Realtors – and salespeople in general – 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 [Prus, 1989].

Nonverbal behaviour plays an especially important role in such social dialogue, as evidenced by the fact that most important business meetings are still conducted face-to-face rather than on the phone. This intuition is backed up by empirical research; several studies have found that the additional nonverbal cues provided by video-mediated communication do not effect performance in task-oriented interactions, but in interactions of a more social nature, such as getting acquainted or negotiation, video is superior [Whittaker and O'Conaill, 1997]. These studies have found that for social tasks, interactions were more personalized, less argumentative and more polite when conducted via video-mediated communication, that participants believed video-mediated (and face-to-face) communication was superior, and that groups conversing using video-mediated communication tended to like each other more, compared to audio-only interactions.

Together, these findings indicate that if we are to develop computer agents capable of performing as well as humans on tasks such as real estate sales then, in addition to task goals such reliable and efficient information delivery, they must have the appropriate social competencies designed into them. Further, since these competencies include the use of nonverbal behaviour for conveying communicative and social cues, then our agents must have the capability of producing and recognizing nonverbal cues in simulations of face-to-face interactions. We call agents with such capabilities "Embodied Conversational Agents" or "ECAs."

The current chapter extends previous work which demonstrated that social dialogue can have a significant impact on a user's trust of an ECA [Bickmore and Cassell, 2001], by investigating whether these results hold in the absence of nonverbal cues. We present the results of a study designed to determine whether the psychological effects of social dialogue – namely to increase trust and associated positive evaluations – vary when the nonverbal cues provided by the embodied conversational agent are removed. In addition to varying medium (voice only vs. embodied) and dialogue style (social dialogue vs. task-only) we also assessed and examined effects due to the user's personality along the introversion/extroversion dimension, since extroversion is one indicator of a person's comfort level with face-to-face interaction.

2. Embodied Conversational Agents

Embodied conversation agents are animated anthropomorphic interface agents that are able to engage a user in real-time, multimodal dialogue, using speech, gesture, gaze, posture, intonation, and other verbal and nonverbal behaviours to emulate the experience of human face-to-face interaction [Cassell et al., 2000c]. The nonverbal channels are important not only for conveying information (redundantly or complementarily with respect to the speech channel), but also for regulating the flow of the conversation. The nonverbal channel is especially crucial for social dialogue, since it can be used to provide such social cues as attentiveness, positive affect, and liking and attraction, and to mark shifts into and out of social activities [Argyle, 1988].

2.1 Functions versus Behaviours

Embodiment provides the possibility for a wide range of behaviours that, when executed in tight synchronization with language, carry out a communicative function. It is important to understand that particular behaviours, such as the raising of the eyebrows, can be employed in a variety of circumstances to produce different communicative effects, and that the same communicative function may be realized through different sets of behaviours. It is therefore clear that any system dealing with conversational modelling has to handle function separately from surface-form or run the risk of being inflexible and insensitive to the natural phases of the conversation. Here we briefly describe some of the fundamental communication categories and their functional sub-parts, along with examples of nonverbal behaviour that contribute to their successful implementation. Table 1.1 shows examples of mappings from communicative function to particular behaviours and is based on previous research on typical North American nonverbal displays, mainly [Chovil, 1991; Duncan, 1974; Kendon, 1980].

Conversation initiation and termination Humans partake in an elaborate ritual when engaging and disengaging in conversations [Kendon, 1980]. For example, people will show their readiness to engage in a conversation by turning towards the potential interlocutors, gazing at them and then exchanging signs of mutual recognition typically involving a smile, eyebrow movement and tossing the head or waving of the arm. Following this initial synchronization stage, or distance salutation, the two people approach each other, sealing their commitment to the conversation through a close salutation such as a handshake accompanied by a ritualistic verbal exchange. The greeting phase ends when the two participants re-orient their bodies, moving away from a face-on orientation to stand at an angle. Terminating a conversation similarly moves through stages, starting with non-verbal cues, such as orientation shifts

Table 1.1. Some examples of conversational functions and their behaviour realization [Cassell et al., 2000b].

Communicative Functions	Communicative Behaviour
Initiation and termination	
Reacting	Short Glance
Inviting Contact	Sustained Glance, Smile
Distance Salutation	Looking, Head Toss/Nod, Raise Eyebrows, Wave, Smile
Close Salutation	Looking, Head Nod, Embrace or Handshake, Smile
Break Away	Glance Around
Farewell	Looking, Head Nod, Wave
Turn-Taking	
Give Turn	Looking, Raise Eyebrows (followed by silence)
Wanting Turn	Raise Hands into gesture space
Take Turn	Glance Away, Start talking
Feedback	
Request Feedback	Looking, Raise Eyebrows
Give Feedback	Looking, Head Nod

or glances away and cumulating in the verbal exchange of farewells and the breaking of mutual gaze.

Conversational turn-taking and interruption Interlocutors do not normally talk at the same time, thus imposing a turn-taking sequence on the conversation. The protocols involved in floor management – determining whose turn it is and when the turn should be given to the listener – involve many factors including gaze and intonation [Duncan, 1974]. In addition, listeners can interrupt a speaker not only with voice, but also by gesturing to indicate that they want the turn.

Content elaboration and emphasis Gestures can convey information about the content of the conversation in ways for which the hands are uniquely suited. For example, the two hands can better indicate simultaneity and spatial relationships than the voice or other channels. Probably the most commonly thought of use of the body in conversation is the pointing (deictic) gesture, possibly accounting for the fact that it is also the most commonly implemented for the bodies of animated interface agents. In fact, however, most conversations don't involve many deictic gestures [McNeill, 1992] unless the interlocutors are discussing a shared task that is currently present. Other conversational gestures also convey semantic and pragmatic information. Beat gestures are small, rhythmic baton like movements of the hands that do not change in form with the content of the accompanying speech. They serve a pragmatic func-

tion, conveying information about what is "new" in the speaker's discourse. Iconic and metaphoric gestures convey some features of the action or event being described. They can be redundant or complementary relative to the speech channel, and thus can convey additional information or provide robustness or emphasis with respect to what is being said. Whereas iconics convey information about spatial relationships or concepts, metaphorics represent concepts which have no physical form, such as a sweeping gesture accompanying "the property title is free and clear."

Feedback and error correction During conversation, speakers can nonverbally request feedback from listeners through gaze and raised eyebrows and listeners can provide feedback through head nods and paraverbals ("uh-huh", "mmm", etc.) if the speaker is understood, or a confused facial expression or lack of positive feedback if not. The listener can also ask clarifying questions if they did not hear or understand something the speaker said.

2.2 Interactional versus Propositional Behaviours

The mapping from form (behaviour) to conversational function relies on a fundamental division of conversational goals: contributions to a conversation can be propositional and interactional. Propositional information corresponds to the content of the conversation. This includes meaningful speech as well as hand gestures and intonation used to complement or elaborate upon the speech content (gestures that indicate the size in the sentence "it was this big" or rising intonation that indicates a question with the sentence "you went to the store"). Interactional information consists of the cues that regulate conversational process and includes a range of nonverbal behaviours (quick head nods to indicate that one is following) as well as regulatory speech ("huh?", "Uh-huh"). This theoretical stance allows us to examine the role of embodiment not just in task- but also process-related behaviours such as social dialogue [Cassell et al., 2000b].

2.3 REA

Our platform for conducting research into embodied conversational agents is the REA system, developed in the Gesture and Narrative Language Group at the MIT Media Lab [Cassell et al., 2000a]. REA is an embodied, multi-modal real-time conversational interface agent which implements the conversational protocols described above in order to make interactions as natural as face-to-face conversation with another person. In the current task domain, REA acts as a real estate salesperson, answering user questions about properties in her database and showing users around the virtual houses (Figure 1.1).

Figure 1.1. User interacting with REA.

REA has a fully articulated graphical body, can sense the user passively through cameras and audio input, and is capable of speech with intonation, facial display, and gestural output. The system currently consists of a large projection screen on which REA is displayed and which the user stands in front of. Two cameras mounted on top of the projection screen track the user's head and hand positions in space. Users wear a microphone for capturing speech input. A single SGI Octane computer runs the graphics and conversation engine of REA, while several other computers manage the speech recognition and generation and image processing.

REA is able to conduct a conversation describing the features of the task domain while also responding to the users' verbal and non-verbal input. When the user makes cues typically associated with turn taking behaviour 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 error correction when she misunderstands what the user says, and can generate combined voice, facial expression and gestural output. REA's responses are generated by an incremental natural language generation engine based on [Stone and Doran, 1997] that has been extended to synthesize redundant and complementary gestures synchronized with speech output [Cassell et al., 2000b]. A simple discourse

model is used for determining which speech acts users are engaging in, and resolving and generating anaphoric references.

3. Social Dialogue

Social dialogue is talk in which interpersonal goals are foregrounded and task goals – if existent – are backgrounded. One of the most familiar contexts in which social dialogue occurs is in human social encounters between individuals who have never met or are unfamiliar with each other. In these situations conversation is usually initiated by "small talk" in which "light" conversation is made about neutral topics (e.g., weather, aspects of the interlocutor's physical environment) or in which personal experiences, preferences, and opinions are shared [Laver, 1981]. Even in business or sales meetings, it is customary (at least in American culture) to begin with some amount of small talk before "getting down to business".

3.1 The Functions of Social Dialogue

The purpose of small talk is primarily to build rapport and trust among the interlocutors, provide time for them to "size each other up", establish an interactional style, and to allow them to establish their reputations [Dunbar, 1996]. Although small talk is most noticeable at the margins of conversational encounters, it can be used at various points in the interaction to continue to build rapport and trust [Cheepen, 1988], and in real estate sales, a good agent will continue to focus on building rapport throughout the relationship with a buyer [Garros, 1999].

Small talk has received sporadic treatment in the linguistics literature, starting with the seminal work of Malinowski who defined "phatic communion" as "a type of speech in which ties of union are created by a mere exchange of words". Small talk is the language used in free, aimless social intercourse, which occurs when people are relaxing or when they are accompanying "some manual work by gossip quite unconnected with what they are doing" [Malinowski, 1923]. Jacobson also included a "phatic function" in his well-known conduit model of communication, that function being focused on the regulation of the conduit itself (as opposed to the message, sender, or receiver) [Jakobson, 1960]. More recent work has further characterized small talk by describing the contexts in which it occurs, topics typically used, and even grammars which define its surface form in certain domains [Cheepen, 1988; Laver, 1975; Schneider, 1988]. In addition, degree of "phaticity" has been proposed as a persistent goal which governs the degree of politeness in all utterances a speaker makes, including task-oriented ones [Coupland et al., 1992].

3.2 The Relationship between Social Dialogue and Trust

Figure 1.2 outlines the relationship between small talk and trust. REA's dialogue planner represents the relationship between her and the user using a multi-dimensional model of interpersonal relationship based on [Svennevig, 1999]:

familiarity 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 [Altman and Taylor, 1973].

solidarity is defined as "like-mindedness" or having similar behaviour 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 and Levinson, 1978]. There is a correlation between frequency of contact and solidarity, but it is not necessarily a causal relation [Brown and Levinson, 1978; Brown and Gilman, 1972].

affect 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 and Gilman, 1989].

The mechanisms by which small talk are hypothesized to effect trust include facework, coordination, building common ground, and reciprocal appreciation.

Facework The notion of "face" is "the positive social value a person effectively claims for himself by the social role others assume he has taken during a particular contact" [Goffman, 1967]. Interactants maintain face by having their social role 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. Small talk avoids *face threat* (and therefore maintains *solidarity*) by keeping conversation at a safe level of depth.

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]. Small talk increases *coordination* between the two participants by allowing them to synchronize short units of talk and nonverbal acknowledgement (and therefore leads to increased liking and positive *affect*).

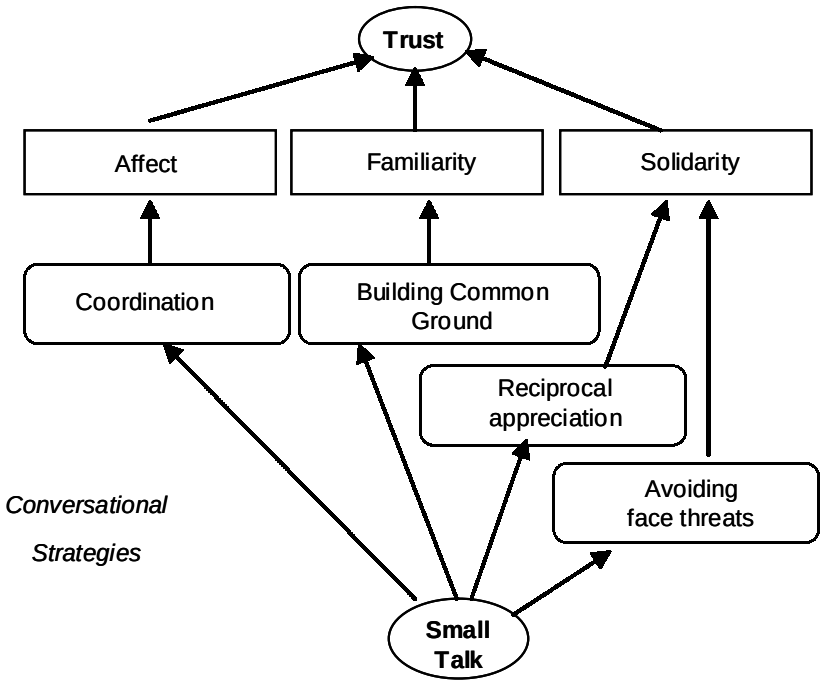

Figure 1.2. How small talk effects trust [Cassell and Bickmore, 2003].

Building common ground 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 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 personal 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. Small talk establishes *common ground* (and therefore increases *familiarity*) by discussing topics that are clearly in the context of utterance.

Reciprocal appreciation In small talk, demonstrating appreciation for and agreement with the contributions of one's interlocutor is obligatory. Performing this aspect of the small talk ritual increases solidarity by showing mutual agreement on the topics discussed.

3.3 Nonverbal Behaviour in Social Dialogue

According to Argyle, nonverbal behaviour is used to express emotions, to communicate interpersonal attitudes, to accompany and support speech, for self presentation, and to engage in rituals such as greetings [Argyle, 1988]. Of these, coverbal and emotional display behaviours have received the most attention in the literature on embodied conversational agents and facial and character animation in general, e.g. [Cassell et al., 2000c]. Next to these, the most important use of nonverbal behaviour in social dialogue is the display of interpersonal attitude [Argyle, 1988]. The display of positive or negative attitude can greatly influence whether we approach someone or not and our initial perceptions of them if we do.

The most consistent finding in this area is that the use of nonverbal "immediacy behaviours" – close conversational distance, direct body and facial orientation, forward lean, increased and direct gaze, smiling, pleasant facial expressions and facial animation in general, nodding, frequent gesturing and postural openness – projects liking for the other and engagement in the interaction, and is correlated with increased solidarity [Argyle, 1988; Richmond and McCroskey, 1995].

Other nonverbal aspects of "warmth" include kinesic behaviours such as head tilts, bodily relaxation, lack of random movement, open body positions, and postural mirroring and vocalic behaviours such as more variation in pitch, amplitude, duration and tempo, reinforcing interjections such as "uh-huh" and "mm-hmmm", greater fluency, warmth, pleasantness, expressiveness, and clarity and smoother turn-taking [Andersen and Guerrero, 1998].

In summary, nonverbal behaviour plays an important role in all face-to-face interaction – both conveying redundant and complementary propositional information (with respect to speech) and regulating the structure of the interaction. In social dialogue, however, it provides the additional, and crucial, function, of conveying attitudinal information about the nature of the relationship between the interactants.

4. Related Work

4.1 Related Work on Embodied Conversational Agents

Work on the development of ECAs, as a distinct field of development, is best summarized in [Cassell et al., 2000c]. The current study is based on the REA ECA (see Figure 1.1), a simulated real-estate agent, who uses vision-based gesture recognition, speech recognition, discourse planning, sentence and gesture planning, speech synthesis and animation of a 3D body [Cassell et al., 1999]. Some of the other major systems developed to date are Steve [Rickel and Johnson, 1998], the DFKI Persona [André et al., 1996], Olga [Beskow and

McGlashan, 1997], and pedagogical agents developed by Lester et al. [1999]. Sidner and Dzikovska [2005] report progress on a robotic ECA that performs hosting activities, with a special emphasis on "engagement" – an interactional behaviour whose purpose is to establish and maintain the connection between interlocutors during a conversation. These systems vary in their linguistic generativity, input modalities, and task domains, but all aim to engage the user in natural, embodied conversation.

Little work has been done on modelling social dialogue with ECAs. The August system is an ECA kiosk designed to give information about local restaurants and other facilities. In an experiment to characterize the kinds of things that people would say to such an agent, over 10,000 utterances from over 2,500 users were collected. It was found that most people tried to socialize with the agent, with approximately 1/3 of all recorded utterances classified as social in nature [Gustafson et al., 1999].

4.2 Related Studies on Embodied Conversational Agents

Koda and Maes [1996] and Takeuchi and Naito [1995] studied interfaces with static or animated faces, and found that users rated them to be more engaging and entertaining than functionally equivalent interfaces without a face. Kiesler and Sproull [1997] found that users were more likely to be cooperative with an interface agent when it had a human face (vs. a dog or cartoon dog).

André, Rist and Muller found that users rated their animated presentation agent ("PPP Persona") as more entertaining and helpful than an equivalent interface without the agent [André et al., 1998]. However, there was no difference in actual performance (comprehension and recall of presented material) in interfaces with the agent vs. interfaces without it. In another study involving this agent, van Mulken, André and Muller found that when the quality of advice provided by an agent was high, subjects actually reported trusting a text-based agent more than either their ECA or a video-based agent (when the quality of advice was low there were no significant differences in trust ratings between agents) [van Mulken et al., 1999].

In a user study of the Gandalf system [Cassell et al., 1999], users rated the smoothness of the interaction and the agent's language skills significantly higher under test conditions in which Gandalf utilized limited conversational behaviour (gaze, turn-taking and beat gesture) than when these behaviours were disabled.

In terms of social behaviours, Sproull et al. [1997] showed that subjects rated a female embodied interface significantly lower in sociability and gave it a significantly more negative social evaluation compared to a text-only interface. Subjects also reported being less relaxed and assured when interacting with the embodied interface than when interacting with the text interface. Fi-

nally, they gave themselves significantly higher scores on social desirability scales, but disclosed less (wrote significantly less and skipped more questions in response to queries by the interface) when interacting with an embodied interface vs. a text-only interface. Men were found to disclose more in the embodied condition and women disclosed more in the text-only condition.

Most of these evaluations have tried to address whether embodiment of a system is useful at all, by including or not including an animated figure. In their survey of user studies on embodied agents, Dehn and van Mulken conclude that there is no "persona effect", that is a general advantage of an interface with an animated agent over one without an animated agent [Dehn and van Mulken, 2000]. However, they believe that lack of evidence and inconsistencies in the studies performed to date may be attributable to methodological shortcomings and variations in the kinds of animations used, the kinds of comparisons made (control conditions), the specific measures used for the dependent variables, and the task and context of the interaction.

4.3 Related Studies on Mediated Communication

Several studies have shown that people speak differently to a computer than another person, even though there are typically no differences in task outcomes in these evaluations. Hauptmann and Rudnicky [1988] performed one of the first studies in this area. They asked subjects to carry out a simple information-gathering task through a (simulated) natural language speech interface, and compared this with speech to a co-present human in the same task. They found that speech to the simulated computer system was telegraphic and formal, approximating a command language. In particular, when speaking to what they believed to be a computer, subject's utterances used a small vocabulary, often sounding like system commands, with very few task-unrelated utterances, and fewer filled pauses and other disfluencies.

These results were extended in research conducted by Oviatt [Oviatt, 1995; Oviatt and Adams, 2000; Oviatt, 1998], in which she found that speech to a computer system was characterized by a low rate of disfluencies relative to speech to a co-present human. She also noted that visual feedback has an effect on disfluency: telephone calls have a higher rate of disfluency than co-present dialogue. From these results, it seems that people speak more carefully and less naturally when interacting with a computer.

Boyle et al. [1994] compared pairs of subjects working on a map-based task who were visible to each other with pairs of subjects who were co-present but could not see each other. Although no performance difference was found between the two conditions, when subjects could not see one another, they compensated by giving more verbal feedback and using longer utterances. Their conversation was found to be less smooth than that between mutually visible

partners, indicated by more interruptions, and less efficient, as more turns were required to complete the task. The researchers concluded that visual feedback improves the smoothness and efficiency of the interaction, but that we have devices to compensate for this when visibility is restricted.

Daly-Jones et al. [1998] also failed to find any difference in performance between video-mediated and audio-mediated conversations, although they did find differences in the quality of the interactions (e.g., more explicit questions in audio-only condition).

Whittaker and O'Conaill [1997] survey the results of several studies which compared video-mediated communication with audio-only communication and concluded that the visual channel does not significantly impact performance outcomes in task-oriented collaborations, although it does affect social and affective dimensions of communication. Comparing video-mediated communication to face-to-face and audio-only conversations, they also found that speakers used more formal turn-taking techniques in the video condition even though users reported that they perceived many benefits to video conferencing relative to the audio-only mode.

In a series of studies on the effects of different media and activities on trust, Zheng, Veinott et al. have demonstrated that social interaction, even if carried out online, significantly increases people's trust in each other [Zheng et al., 2002]. Similarly, Bos et al. [2002] demonstrated that richer media – such as face-to-face, video-, and audio-mediated communication – leads to higher trust levels than media with lower bandwidth such as text chat.

Finally, a number of studies have been done comparing face-to-face conversations with conversations on the phone [Rutter, 1987]. These studies find that, in general, there is more cooperation and trust in face-to-face interaction. One study found that audio-only communication encouraged negotiators to behave impersonally, to ignore the subtleties of self-presentation, and to concentrate primarily on pursuing victory for their side. Other studies found similar gains in cooperation among subjects playing prisoner's dilemma face-to-face compared to playing it over the phone. Face-to-face interactions are also less formal and more spontaneous than conversations on the phone. One study found that face-to-face discussions were more protracted and wide-ranging while subjects communicating via audio-only kept much more to the specific issues on the agenda (the study also found that when the topics were more wide-ranging, changes in attitude among the participants was more likely to occur). Although several studies found increases in favourable impressions of interactants in face-to-face conversation relative to audio-only, these effects have not been consistently validated.

4.4 Trait-Based Variation in User Responses

Several studies have shown that users react differently 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 [Nass and Gong, 2000; Reeves and Nass, 1996]. Resnick and Lammers showed that in order to change user behaviour 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 and 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 and Reeves, 2000].

In our earlier study on the effects of social dialogue on trust in ECA interactions, we found that social dialogue significantly increased trust for extraverts, while it made no significant difference for introverts [Cassell and Bickmore, 2003]. In light of the studies summarized here, the question that remains is whether these effects continue to hold if the nonverbal cues provided by the ECA are removed.

5. Social Dialogue in REA

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 relational model outlined above. An overview of the planner is provided here; details of its implementation can be found in Cassell and Bickmore [2003].

5.1 Planning Model

Given that many of the goals in a relational conversational strategy are non-discrete (e.g., minimize face threat), and that trade-offs among multiple goals have to be achieved at any given time, we have moved away from static world discourse planning, and are using an activation network-based approach based on Maes' *Do the Right Thing* architecture [Maes, 1989]. This architecture provides the capability to transition smoothly from deliberative, planned behaviour to opportunistic, reactive behaviour, and is able to pursue multiple, non-discrete goals. In our implementation each node in the network represents a conversational move that REA can make.

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

Closeness REA continually assesses her "interpersonal" closeness with the user, which is a composite representing depth of familiarity and solidarity, modelled as a scalar quantity. Each conversational topic has a predefined, pre-requisite closeness that must be achieved before REA can introduce the topic. Given this, the system can plan to perform small talk in order to "grease the tracks" for task talk, especially about sensitive topics like finance.

Topic REA keeps track of the current and past conversational topics. Conversational moves which stay within topic (maintain topic coherence) are given preference over those which do not. In addition, REA can plan to execute a sequence of moves which gradually transition the topic from its current state to one that REA wants to talk about (e.g., from talk about the weather, to talk about Boston weather, to talk about Boston real estate).

Relevance REA maintains a list of topics that she thinks the user knows about, and the discourse planner prefers moves which involve topics in this list. The list is initialized to things that anyone talking to REA would know about – such as the weather outside, Cambridge, MIT, or the laboratory that REA lives in.

Task goals REA has a list of prioritized goals to find out about the user's housing needs in the initial interview. Conversational moves which directly work towards satisfying these goals (such as asking interview questions) are preferred.

Logical preconditions Conversational moves have logical preconditions (e.g., it makes no sense for REA to ask users what their major is until she has

established that they are students), and are not selected for execution until all of their preconditions are satisfied.

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

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

At each step in the conversation in which REA has the floor (as tracked by a conversational state machine in REA's Reaction Module [Cassell et al., 2000a], the discourse planner is consulted for the next conversational move to initiate. At this point, activation values are incrementally propagated through the network (following [Maes, 1989]) until a move is selected whose preconditions are satisfied and whose activation value is over a specified threshold.

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).

5.2 Interactional Behaviour during Social Dialogue

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]. Prior to producing lengthy utterances, REA gazes away briefly before she starts her turn, partly as a turn-taking and floor-holding move and partly to mask the processing delays in generating long utterances. Finally, REA smiles as soon as she detects that users have started their speaking turns (using audio thresholding of the user's speech).

5.3 Example Interaction

An interview between REA and a user typically proceeds as shown in the following dialogue. (User responses are only shown in positions in which they

effect the selection of subsequent moves; REA contributions are her actual utterances.)

1 That microphone is terrible, I hate using those things.

2 Sorry about my voice, this is some engineer's idea of natural sounding.

3 Are you one of our sponsors? *User: Yes*

4 Were you at our last sponsor meetings?

5 I got so exhausted at the last sponsor meeting I think I was starting to lose my voice by the end.

6 So, where would you like to live?

7 How many bedrooms do you need?

8 Do you need access to the subway?

9 Is one bath enough?

10 You know, Boston is certainly more expensive than it used to be.

11 Anyway, what can you afford?

12 What kind of down payment can you make?

13 Let me see what I have available.

Dialogue 1. "Small Talk" REA.

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

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

1 So, where would you like to live?

2 What can you afford?

3 What kind of down payment can you make?

4 How many bedrooms do you need?

5 Do you need access to the subway?

6 Is one bath enough?

7 Let me see what I have available.

Dialogue 2. "Task-only REA".

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

6. A Study Comparing ECA Social Dialogue with Audio-Only Social Dialogue

The dialogue model presented above produces a reasonable facsimile of the social dialogue observed in service encounters such as real estate sales. But, does small talk produced by an ECA in a sales encounter actually build trust and solidarity with users? And, does nonverbal behaviour play the same critical role in human-ECA social dialogue as it appears to play in human-human social interactions?

In order to answer these questions, we conducted an empirical study in which subjects were interviewed by REA about their housing needs, shown two "virtual" apartments, and then asked to submit a bid on one of them. For the purpose of the experiment, REA was controlled by a human wizard and followed scripts identical to the output of the planner (but faster, and not dependent on automatic speech recognition or computational vision). Users interacted with one of two versions of REA which were identical except that one had only task-oriented dialogue (TASK condition) while the other also included the social dialogue designed to avoid face threat, and increase trust (SOCIAL condition). A second manipulation involved varying whether subjects interacted with the fully embodied REA – appearing in front of the virtual apartments as a life-sized character (EMBODIED condition) – or viewed only the virtual apartments while talking with REA over a telephone. Together these variables provided a 2x2 experimental design: SOCIAL vs. TASK and EMBODIED vs. PHONE.

Our hypotheses follow from the literature on small talk and on trust among humans. We expected subjects in the SOCIAL condition to trust REA more, feel closer to REA, like her more, and feel that they understood each other more

than in the TASK condition. We also expected users to think the interaction was more natural, lifelike, and comfortable in the SOCIAL condition. Finally, we expected users to be willing to pay REA more for an apartment in the SOCIAL condition, given the hypothesized increase in trust. We also expected all of these SOCIAL effects to be amplified in the EMBODIED condition relative to the PHONE-only condition.

6.1 Experimental Methods

This was a multivariate, multiple-factor, between-subjects experimental design, involving 58 subjects (69% male and 31% female).

6.1.1 Apparatus. One wall of the experiment room was a rear-projection screen. In the EMBODIED condition REA appeared life-sized on the screen, in front of the 3D virtual apartments she showed, and her synthetic voice was played through two speakers on the floor in front of the screen. In the PHONE condition only the 3D virtual apartments were displayed and subjects interacted with REA over an ordinary telephone placed on a table in front of the screen.

For the purpose of this experiment, REA was controlled via a wizard-of-oz setup on another computer positioned behind the projection screen. The interaction script included verbal and nonverbal behaviour 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 was 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 were comprised of a linear sequence of utterances (statements and questions) that would be made by REA in a given interaction: there was no branching or variability in content beyond the three pieces of information described above. This helped ensure that all subjects received the same intervention regardless of what they said in response to any given question by REA. Subject-initiated utterances were responded to with either backchannel feedback (e.g., “Really?”) for statements or “I don’t know” for questions, followed by an immediate return to the script.

The scripts for the TASK and SOCIAL conditions were identical, except that the SOCIAL script had additional small talk utterances added to it, as

described in [Bickmore and Cassell, 2001]. 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. In both conditions subjects were told that they could talk to REA "just like you would to another person".

6.1.2 Measures. Subjective evaluations of REA – including how friendly, credible, lifelike, warm, competent, reliable, efficient, informed, knowledgeable and intelligent she was – were measured by single items on nine-point Likert scales. Evaluations of the interaction – including how tedious, involving, enjoyable, natural, satisfying, fun, engaging, comfortable and successful it was – were also measured on nine-point Likert scales. Evaluation of how well subjects felt they knew REA, how well she knew and understood them and how close they felt to her were measured in the same manner. All scales were adapted from previous research on user responses to personality types in embodied conversational agents [Moon and Nass, 1996].

Liking of REA was an index composed of three items – how likeable and pleasant REA was and how much subjects liked her – measured items on nine-point Likert scales (Cronbach's alpha =.87).

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.

Trust was measured by a standardized trust scale [Wheeless and Grotz, 1977] (alpha =.93). Although trust is sometimes measured behaviourally using a Prisoner's Dilemma game [Zheng et al., 2002], we felt that our experimental protocol was already too long and that game-playing did not fit well into the real estate scenario.

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

Extrovertedness was an index composed of seven Wiggins [Wiggins, 1979] extrovert adjective items: Cheerful, Enthusiastic, Extroverted, Jovial, Outgo-

ing, and Perky. It was used for assessment of the subject's personality (alpha =.87).

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's personality (alpha =.84). Note that these personality scales were administered on the post-test questionnaire. For the purposes of this experiment, therefore, subjects who scored over the mean on introversion-extroversion were said to be extroverts, while those who scored under the mean were said to be introverts.

6.1.3 Behavioural measures. Rates of speech disfluency (as defined in [Oviatt, 1995]) and utterance length were coded from the video data.

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 *ACTIVE* (above the mean on number of user-initiated utterances). To our surprise, these measures turned out to be independent of introversion/extroversion (Pearson r=0.042), and to not be predicted by these latter variables.

6.2 Results

Full factorial single measure ANOVAs were run, with SOCIALITY (Task vs. Social), PERSONALITY OF SUBJECT (Introvert vs. Extrovert), MEDIUM (Phone vs. Embodied) and INITIATION (Active vs. Passive) as independent variables.

6.2.1 Subjective assessments of REA. In looking at the questionnaire data, our first impression is that subjects seemed to feel more comfortable interacting with REA over the phone than face-to-face. Thus, subjects in the phone condition felt that they knew REA better ($F=5.02$; $p<.05$), liked her more ($F=4.70$; $p<.05$), felt closer to her ($F=13.37$; $p<.001$), felt more comfortable with the interaction ($F=3.59$; $p<.07$), and thought REA was more friendly ($F=8.65$;$p<.005$), warm ($F=6.72$; $p<.05$), informed ($F=5.73$; $p<.05$), and knowledgeable ($F=3.86$; $p<.06$) than those in the embodied condition.

However, in the remainder of the results section, as we look more closely at different users, different kinds of dialogue styles, and users' actual behaviour, a more complicated picture emerges. Subjects felt that REA knew them ($F=3.95$; $p<.06$) and understood them ($F=7.13$; $p<.05$) better when she used task-only dialogue face-to-face; these trends were reversed for phone-based interactions. Task-only dialogue was more fun ($F=3.36$; $p<.08$) and less tedious ($F=8.77$;

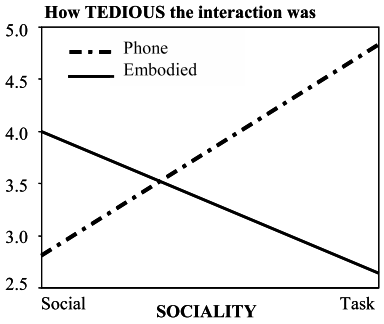

Figure 1.3. Ratings of TEDIOUS.

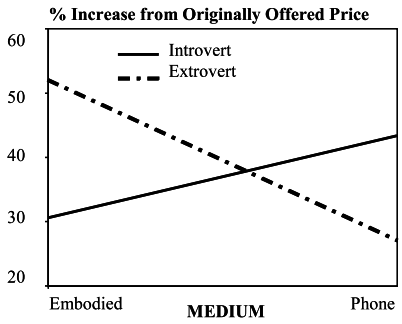

Figure 1.4. Amount subjects were willing to pay.

p<.005; see Figure 1.3) when embodied, while social dialogue was more fun and less tedious on the phone. That is, subjects preferred to interact, and felt better understood, face-to-face when it was a question of simply "getting down to business," and preferred to interact, and felt better understood, by phone when the dialogue included social chit-chat.

These results may be telling us that REA's nonverbal behaviour inadvertently projected an unfriendly, introverted personality that was especially inappropriate for social dialogue. REA's model of non-verbal behaviour, at the time of this experiment, was limited to those behaviours linked to the discourse context. Thus, REA's smiles were limited to those related to the ends of turns, and she did not have a model of immediacy or other nonverbal cues for liking and warmth typical of social interaction [Argyle, 1988]. According to Whittaker and O'Connail [1993], non-verbal information is especially crucial in interactions involving affective cues, such as negotiation or relational dialogue, and less important in purely problem-solving tasks. This interpretation of the results is backed up by comments such as this response from a subject in the face-to-face social condition:

> The only problem was how she would respond. She would pause then just say "OK", or "Yes". Also when she looked to the side and then back before saying something was a little bit unnatural.

This may explain why subjects preferred task interactions face-to-face, while on the phone REA's social dialogue had its intended effect of making subjects feel that they knew REA better, that she understood them better, and that the experience was more fun and less tedious.

In our earlier study, looking only at an embodied interface, we reported that extroverts trusted the system more when it engaged in small talk, while introverts were not affected by the use of small talk [Bickmore and Cassell, 2001]. In the current study, these results were re-confirmed, but only in the embodied

interaction; that is, a three-way interaction between SOCIALITY, PERSONALITY and MEDIUM ($F=3.96$; $p<.06$) indicated that extroverts trusted REA more when she used social dialogue in embodied interactions, but there was essentially no effect of user's personality and social dialogue on trust in phone interactions. Further analysis of the data indicates that this result derives from the substantial difference between introverts and extroverts in the face-to-face task-only condition. Introverts trusted REA significantly more in the face-to-face task-only condition than in the other conditions ($p<.03$), while extroverts trusted her significantly less in this condition than in the other conditions ($p.<01$).

In light of these new observations, our earlier results indicating that social dialogue leads to increased trust (for extroverts at least) needs to be revised. This further analysis indicates that the effects we observed may be due to the attraction of a computer displaying similar personality characteristics, rather than the process of trust-building. That is, in the face-to-face, task-only condition, both verbal and nonverbal channels appear to have inadvertently indicated that REA was an introvert (also supported by the comments that REA's gaze-away behaviour was too frequent, an indication of introversion [Wilson, 1977]), and in this condition we find the introverts trusting more, and extroverts trusting less. In all other conditions, the personality cues are either conflicting (a mismatch between verbal and nonverbal behaviour has been demonstrated to be disconcerting to users [Nass and Gong, 2000]) or only one channel of cues is available (i.e. on the phone), yielding trust ratings that are close to the overall mean.

There was, nevertheless, a preference by extroverts for social dialogue as demonstrated by the fact that, overall, extroverts liked REA more when she used social dialogue, while introverts liked her more when she only talked about the task ($F=8.09$; $p<.01$).

Passive subjects felt more comfortable interacting with REA than active subjects did, regardless of whether the interaction was face-to-face or on the phone, or whether REA used social dialogue or not. Passive subjects said that they enjoyed the interaction more ($F=4.47$; $p<.05$), felt it was more successful ($F=6.04$; $p<.05$) and liked REA more ($F=3.24$; $p<.08$), and that REA was more intelligent ($F=3.40$; $p<.08$), and knew them better ($F=3.42$; $p<.08$) than active subjects. These differences may be explained by the fixed-initiative dialogue model used in the WOZ script. REA's interaction was designed for passive users – there was very little capability in the interaction script to respond to unanticipated user questions or statements – and user initiation attempts were typically met with uncooperative system responses or interruptions. But, given the choice between phone and face-to-face, passive users preferred to interact with REA face-to-face: they rated her as more friendly ($F=3.56$; $p<.07$) and informed ($F=6.30$; $p<.05$) in this condition. Passive users also found the phone

to be more tedious, while active users also found the phone to be less tedious ($F=5.15$; $p<.05$). Active users may have found the face-to-face condition particularly frustrating since processing delays may have led to the perception that the floor was open (inviting an initiation attempt), when in fact the wizard had already instructed REA to produce her next utterance.

However, when interacting on the phone, active users differed from passive users in that active users felt she was more reliable when using social dialogue and passive users felt she was more reliable when using task-only dialogue. When interacting face-to-face with REA, there was no such distinction between active and passive users ($F=4.67$; $p<.05$).

6.2.2 Effects on task measure. One of the most tantalizing results obtained is that extroverts were willing to pay more for the same apartment in the embodied condition, while introverts were willing to pay more over the phone ($F=3.41$; $p<.08$), as shown in Figure 1.4.

While potentially very significant, this finding is a little difficult to explain, especially given that trust did not seem to play a role in the evaluation. Perhaps, since we asked our subjects to play the role of someone looking for an apartment, and given that the apartments displayed were cartoon renditions, the subjects may not have felt personally invested in the outcome, and thus may have been more likely to be persuaded by associative factors like the perceived liking and credibility of REA. In fact, trust has been shown to not play a role in persuasion when "peripheral route" decisions are made, which is the case when the outcome is not of personal significance [Petty and Wegener, 1998]. Further, extroverts are not only more sociable, but more impulsive than introverts [Wilson, 1977], and impulse buying is governed primarily by novelty [Onkvisit and Shaw, 1994]. Extroverts did rate face-to-face interaction as more engaging than phone-based interaction (though not at a level of statistical significance), while introverts rated phone-based interactions as more engaging, providing some support for this explanation. It is also possible that this measure tells us more about subjects' assessment of the house than of the realtor. In future experiments we may ask more directly whether the subject perceived the realtor to be asking a fair price. Perception of fairness of a price may be more linked to trust than is actual price demanded for a property.

6.2.3 Gender effects. Women felt that REA was more efficient ($F=5.61$; $p<.05$) and reliable ($F=4.99$; $p<.05$) in the embodied condition than when interacting with her over the phone, while men felt that she was more efficient and reliable by phone. Of course, REA has a female body and a female voice and so in order to have a clearer picture of the meaning of these results, a similar study would need to be carried out with a male realtor.

Table 1.2. Speech disfluencies per 100 words.

	Embodied	*Phone*	*Overall*
Disfluencies	4.83	6.73	5.83

Table 1.3. Speech disfluencies per 100 Words for different types of human-human and simulated human-computer interactions (adapted from Oviatt [Oviatt, 1995]).

Human-human speech	
Two-person telephone call	8.83
Two-person face-to-face dialogue	5.5
Human-computer speech	
Unconstrained computer interaction	1.80
Structured computer interaction	0.83

6.2.4 Effects on behavioural measures. Although subjects' beliefs about REA and about the interaction are important, it is at least equally important to look at how subjects *act*, independent of their conscious beliefs.

In this context we examined subjects' disfluencies when speaking with REA. Remember that disfluency can be a measure of naturalness – human-human conversation demonstrates *more* disfluency than does human-computer communication [Oviatt, 1995]. The rates of speech disfluencies (per 100 words) are shown in Table 1.2. Comparing these to results from previous studies (see Table 1.3) indicates that interactions with REA were more similar to human-human conversation than to human-computer interaction. When asked if he was interacting with a computer or a person, one subject replied "A computer-person I guess. It was a lot like a human."

There were no significant differences in utterance length (MLU) across any of the conditions.

Strikingly, the behavioural measures indicate that, with respect to speech disfluency rates, talking to REA is more like talking to a person than talking to a computer.

Once again, there were significant effects of MEDIUM, SOCIALITY and PERSONALITY on disfluency rate ($F=7.09$; $p<.05$), such that disfluency rates were higher in TASK than SOCIAL, higher overall for INTROVERTs than EXTROVERTs, higher for EXTROVERTs on the PHONE, and higher for INTROVERTs in EMBODIED condition. These effects on disfluency rates are consistent with our conclusion that REA's nonverbal behaviours inadvertently projected an introverted and asocial persona, and with the secondary hypoth-

esis that the primary driver on disfluency is cognitive load, once the length of the utterance is controlled for [Oviatt, 1995]. Given our results, this hypothesis would indicate that social dialogue requires lower cognitive load than task-oriented dialogue, that conversation requires a higher cognitive load on introverts than extraverts, that talking on the phone is more demanding than talking face-to-face for extraverts, and that talking face-to-face is more demanding than talking on the phone for introverts, all of which seem reasonable.

7. Conclusion

The complex results of this study give us hope for the future of embodied conversational agents, but also a clear roadmap for future research. In terms of their behaviour with REA, users demonstrated that they treat conversation with her more like human-human conversation than like human-computer conversation. Their verbal disfluencies are the mark of unplanned speech, of a conversational style. However, in terms of their assessment of her abilities, this did not mean that users saw REA through rose-colored glasses. They were clear about the necessity not only to embody the interaction, but to design every aspect of the embodiment in the service of the same interaction. That is, face-to-face conversations with ECAs must demonstrate the same quick timing of nonverbal behaviours as humans (not an easy task, given the state of the technologies we employ). In addition, the persona and nonverbal behaviour of an ECA must be carefully designed to match the task, a conversational style, and user expectations.

Relative to other studies on social dialogue and interactions with ECAs, this study has taught us a great deal about how to build ECAs capable of social dialogue and the kinds of applications in which this is important to do. As demonstrated in the study of the August ECA, we found that people will readily conduct social dialogue with an ECA in situations in which there is no time pressure to complete a task or in which it is a normal part of the script for similar human-human interactions, and many people actually prefer this style of interaction. Consequently, we are in the process of developing and evaluating an ECA in the area of coaching for health behaviour change [Bickmore, 2002], an area in which social dialogue – and relationship-building behaviours in general – are known to significantly effect task outcomes.

Relative to prior findings in human-human interaction that social dialogue builds trust, our inability to find this effect across all users is likely due to shortcomings in our model, especially in the area of appropriate nonverbal behaviour and lack of uptake of subjects' conversational contributions. Given these shortcomings, similarity attraction effects appear to have overwhelmed the social-dialogue-trust effects, and we observed introverts liking and trusting REA more when she behaved consistently introverted and extraverts liking and

trusting her more when she behaved consistently extraverted. Our conclusion from this is that, adding social dialogue to embodied conversational agents will require a model of social nonverbal behaviour consistent with verbal conversational strategies, before the social dialogue fulfils its trust-enhancing role.

As computers begin to resemble humans, the bar of user expectations is raised: people expect that REA will hold up her end of the conversation, including dealing with interruptions by active users. We have begun to demonstrate the feasibility of embodied interfaces. Now it is time to design ECAs that people wish to spend time with, and that are able to use their bodies for conversational tasks for which human face-to-face interaction is unparalleled, such as social dialogue, initial business meetings, and negotiation.

Acknowledgements

Thanks to Ian Gouldstone, Jennifer Smith and Elisabeth Sylvan for help in conducting the experiment and analyzing data, and to the rest of the Gesture and Narrative Language Group for their help and support.

References

Altman, I. and Taylor, D. (1973). *Social Penetration: The Development of Interpersonal Relationships*. New York: Holt, Rinhart & Winston.

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

André, E., Muller, J., and Rist, T. (1996). The PPP Persona: A Multipurpose Animated Presentation Agent. In *Proceedings of Advanced Visual Interfaces*, pages 245–247, Gubbio, Italy.

André, E., Rist, T., and Muller, J. (1998). Integrating Reactive and Scripted Behaviors in a Life-Like Presentation Agent. In *Proceedings of the Second International Conference on Autonomous Agents*, pages 261–268, Minneapolis, Minnesota, USA.

Argyle, M. (1988). *Bodily Communication*. New York: Methuen & Co. Ltd.

Argyle, M. (1990). The Biological Basis of Rapport. *Psychological Inquiry*, 1:297–300.

Beskow, J. and McGlashan, S. (1997). Olga: A Conversational Agent with Gestures. In André, E., editor, *Proceedings of the IJCAI 1997 Workshop on Animated Interface Agents: Making Them Intelligent*, Nagoya, Japan. San Francisco: Morgan-Kaufmann Publishers.

Bickmore, T. (2002). When Etiquette Really Matters: Relational Agents and Behavior Change. In *Proceedings of AAAI Fall Symposium on Etiquette for Human-Computer Work*, pages 9–10, Falmouth, MA.

Bickmore, T. and Cassell, J. (2001). Relational Agents: A Model and Implementation of Building User Trust. In *Proceedings of CHI 2001*, pages 396–403, Seattle, WA.

Bos, N., Olson, J. S., Gergle, D., Olson, G. M., and Wright, Z. (2002). Effects of Four Computer-Mediated Communications Channels on Trust Development. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 135–140, Minneapolis, Minnesota, USA.

Boyle, E., Anderson, A., and Newlands, A. (1994). The Effects of Visibility in a Cooperative Problem Solving Task. *Language and Speech*, 37(1):1–20.

Brown, P. and Levinson, S. (1978). Universals in Language Usage: Politeness Phenomena. In Goody, E., editor, *Questions and Politeness: Strategies in Social Interaction*, pages 56–289. Cambridge: Cambridge University Press.

Brown, R. and Gilman, A. (1972). The Pronouns of Power and Solidarity. In Giglioli, P., editor, *Language and Social Context*, pages 252–282. Harmondsworth: Penguin.

Brown, R. and Gilman, A. (1989). Politeness Theory and Shakespeare's Four Major Tragedies. *Language in Society*, 18:159–212.

Cassell, J. and Bickmore, T. (2003). Negotiated Collusion: Modeling Social Language and its Relationship Effects in Intelligent Agents. *User Modeling and Adaptive Interfaces*, 13(1-2):89–132.

Cassell, J., Bickmore, T., Billinghurst, M., Campbell, L., Chang, K., Vilhjálmsson, H., and Yan, H. (1999). Embodiment in Conversational Interfaces: Rea. In *Proceedings of ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 520–527, Pittsburgh, PA.

Cassell, J., Bickmore, T., Campbell, L., Vilhjálmsson, H., and Yan, H. (2000a). Human Conversation as a System Framework: Designing Embodied Conversational Agents. In *Embodied Conversational Agents*, pages 29–63. Cambridge, MA: MIT Press.

Cassell, J., Bickmore, T., H, Vilhjálmsson, and Yan, H. (2000b). More Than Just a Pretty Face: Affordances of Embodiment. In *Proceedings of the 5th International Conference on Intelligent User Interfaces*, pages 52–59, New Orleans, Louisiana.

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

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

Chovil, N. (1991). Discourse-Oriented Facial Displays in Conversation. *Research on Language and Social Interaction*, 25(1991/1992):163–194.

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

Coupland, J., Coupland, N., and Robinson, J. D. (1992). How Are You? Negotiating Phatic Communion. *Language in Society*, 21:207–230.

Daly-Jones, O., Monk, A. F., and Watts, L. A. (1998). Some Advantages of Video Conferencing over High-Quality Audio Conferencing: Fluency and Awareness of Attentional Focus. *International Journal of Human-Computer Studies*, 49(1):21–58.

Dehn, D. M. and van Mulken, S. (2000). The Impact of Animated Interface Agents: A Review of Empirical Research. *International Journal of Human-Computer Studies*, 52:1–22.

Dunbar, R. (1996). *Grooming, Gossip, and the Evolution of Language*. Cambridge, MA: Harvard University Press.

Duncan, S. (1974). On the Structure of Speaker-Auditor Interaction during Speaking Turns. *Language in Society*, 3:161–180.

Garros, D. (1999). Real Estate Agent, Home and Hearth Realty. Personal communication.

Goffman, I. (1967). On Face-Work. In *Interaction Ritual: Essays on Face-to-Face Behavior*, pages 5–46. New York: Pantheon.

Gustafson, J., Lindberg, N., and Lundeberg, M. (1999). The August Spoken Dialogue System. In *Proceedings of the Eurospeech 1999 Conference*, pages 1151–1154, Budapest, Hungary.

Hauptmann, A. G. and Rudnicky, A. I. (1988). Talking to Computers: An Empirical Investigation. *International Journal of Man-Machine Studies*, 8(6): 583–604.

Jakobson, R. (1960). Concluding Statement: Linguistics and Poetics. In Sebeok, T., editor, *Style in Language*, pages 351–377. New York: Wiley.

Kendon, A. (1980). *Conducting Interaction: Patterns of Behavior in Focused Encounters*, volume 7. Cambridge: Cambridge University Press.

Kiesler, S. and Sproull, L. (1997). Social Human-Computer Interaction. In Friedman, B., editor, *Human Values and the Design of Computer Technology*, pages 191–199. Stanford, CA: CSLI Publications.

Koda, T. and Maes, P. (1996). Agents with Faces: The Effect of Personification. In *Proceedings of the Fifth IEEE International Workshop on Robot and Human Communication (RO-MAN 1996)*, pages 189–194, Tsukuba, Japan.

Laver, J. (1975). Communicative Functions of Phatic Communion. In Kendon, A., Harris, R., and Key, M., editors, *The Organization of Behavior in Face-to-Face Interaction*, pages 215–238. The Hague: Mouton.

Laver, J. (1981). Linguistic Routines and Politeness in Greeting and Parting. In Coulmas, F., editor, *Conversational Routine*, pages 289–304. The Hague: Mouton.

Lester, J., Stone, B., and Stelling, G. (1999). Lifelike Pedagogical Agents for Mixed-Initiative Problem Solving in Constuctivist Learning Environments. *User Modeling and User-Adapted Interaction*, 9(1-2):1–44.

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 Ogden, C. K. and Richards, I. A., editors, *The Meaning of Meaning*, pages 296–346. Routledge & Kegan Paul.

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. Cambridge: Cambridge University Press.

Moon, Y. and Nass, C. I. (1996). How "Real" are Computer Personalities? Psychological Responses to Personality Types in Human-Computer Interaction. *Communication Research*, 23(6):651–674.

Nass, C. and Gong, L. (2000). Speech Interfaces from an Evolutionary Perspective. *Communications of the ACM*, 43(9):36–43.

Onkvisit, S. and Shaw, J. J. (1994). *Consumer Behavior: Strategy and Analysis*. New York: Macmillan College Publishing Company.

Oviatt, S. (1995). Predicting Spoken Disfluencies during Human-Computer Interaction. *Computer Speech and Language*, 9:19–35.

Oviatt, S. and Adams, B. (2000). Designing and Evaluating Conversational Interfaces with Animated Characters. In Cassell, J., Sullivan, J., Prevost, S., and Churchill, E., editors, *Embodied Conversational Agents*, pages 319–345. Cambridge, MA: MIT Press.

Oviatt, S. L. (1998). User-Centered Modeling for Spoken Language and Multimodal Interfaces. In Maybury, M. T. and Wahlster, W., editors, *Readings in Intelligent User Interfaces*, pages 620–630. San Francisco, CA: Morgan Kaufmann Publishers, Inc.

Petty, R. and Wegener, D. (1998). Attitude Change: Multiple Roles for Persuasion Variables. In Gilbert, D., Fiske, S., and Lindzey, G., editors, *The Handbook of Social Psychology*, pages 323–390. New York: McGraw-Hill.

Prus, R. (1989). *Making Sales: Influence as Interpersonal Accomplishment*. Mewbury Park, CA: Sage.

Reeves, B. and Nass, C. (1996). *The Media Equation*. Cambridge: Cambridge University Press.

Resnick, P. V. and 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.

Richmond, V. and McCroskey, J. (1995). *Immediacy - Nonverbal Behavior in Interpersonal Relations*. Boston: Allyn & Bacon.

Rickel, J. and Johnson, W. L. (1998). Task-Oriented Dialogs with Animated Agents in Virtual Reality. In *Proceedings of the First Workshop on Embodied Conversational Characters*, pages 39–46.

Rickenberg, R. and Reeves, B. (2000). The Effects of Animated Characters on Anxiety, Task Performance, and Evaluations of User Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 49–56, The Hague, Amsterdam.

Rutter, D. R. (1987). *Communicating by Telephone*. New York: Pergamon Press.

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

Sidner, C. and Dzikovska, M. (2005). A First Experiment in Engagement for Human-Robot Interaction in Hosting Activities. In van Kuppevelt, J., Dybkjær, L., and Bernsen, N. O., editors, *Advances in Natural Multimodal Dialogue Systems*. Springer. This volume.

Sproull, L., Subramani, M., Kiesler, S., Walker, J., and Waters, K. (1997). When the Interface is a Face. In Friedman, B., editor, *Human Values and the Design of Computer Technology*, pages 163–190. Stanford, CA: CSLI Publications.

Stone, M. and Doran, C. (1997). Sentence Planning as Description Using Tree-Adjoining Grammar. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics (ACL/EACL 1997)*, pages 198–205, Madrid, Spain.

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

Takeuchi, A. and Naito, T. (1995). Situated Facial Displays: Towards Social Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 450–455, Denver, Colorado.

van Mulken, S., André, E., and Muller, J. (1999). An Empirical Study on the Trustworthiness of Lifelike Interface Agents. In Bullinger, H.-J. and Ziegler, J., editors, *Human-Computer Interaction (Proceedings of HCI-International 1999)*, pages 152–156. Mahwah, NJ: Lawrence Erlbaum Associates.

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

Whittaker, S. and O'Conaill, B. (1993). An Evaluation of Video Mediated Communication. In *Proceedings of Human Factors and Computing Systems (INTERACT/CHI 1993)*, pages 73–74, Amsterdam, The Netherlands.

Whittaker, S. and O'Conaill, B. (1997). The Role of Vision in Face-to-Face and Mediated Communication. In Finn, K., Sellen, A., and Wilbur, S., editors, *Video-Mediated Communication*, pages 23–49. Lawrence Erlbaum Associates, Inc.

Wiggins, J. (1979). A Psychological Taxonomy of Trait-Descriptive Terms. *Journal of Personality and Social Psychology*, 37(3):395–412.

Wilson, G. (1977). Introversion/Extraversion. In Blass, T., editor, *Personality Variables in Social Behavior*, pages 179–218. New York: John Wiley & Sons.

Zheng, J., Veinott, E. S., Bos, N., Olson, J. S., and Olson, G. M. (2002). Trust without Touch: Jumpstarting Long-Distance Trust with Initial Social Activ-

ities. In *Proceedings of the International Conference for Human-Computer Interaction (CHI)*, pages 141–146.