Chapter 3

A VIRTUAL HUMAN DIALOGUE MODEL FOR NON-TEAM INTERACTION

David Traum, William Swartout, Jonathan Gratch and Stacy Marsella
University of Southern California
Marina del Rey, CA, USA
traum@ict.usc.edu, swartout@ict.usc.edu, gratch@ict.usc.edu, marsella@ict.usc.edu

Abstract We describe the dialogue model for the virtual humans developed at the Institute for Creative Technologies at the University of Southern California. The dialogue model contains a rich set of information state and dialogue moves to allow a wide range of behaviour in multimodal, multiparty interaction. We extend this model to enable non-team negotiation, using ideas from social science literature on negotiation and implemented strategies and dialogue moves for this area. We present a virtual human doctor who uses this model to engage in multimodal negotiation dialogue with people from other organisations. The doctor is part of the SASO-ST system, used for training for non-team interactions.

Keywords: Dialogue; negotiation; virtual humans; embodied conversational agents

1. Introduction

Virtual Humans (Rickel and Johnson, 1999b) are autonomous agents who can play the role of people in simulations or games. These agents generally have some or all of the following properties:

- Humanoid body (either a physical robot, or animated body in a virtual environment)
- Cognitive state, including beliefs, desires or goals, intentions, and perhaps other attitudes
- Embeddedness in the real or a virtual world

L. Dybkjær and W. Minker (eds.), Recent Trends in Discourse and Dialogue, 45–67.

- Interactivity with the world (or a virtual world), other virtual humans, and real people, including perception of events and communication, and ability to manipulate the world and/or communicate with outers
- Believable human-like behaviour, including affective reasoning and behaviour

Virtual humans can play an important role in helping train skills of interacting with others who have different beliefs, goals, and styles of behaviour. By building virtual humans that are not just humanoid in appearance and external behaviour, but which also have internal models (including beliefs, goals, plans, and emotions) and ability to reason over these models and formulate appropriate strategies and behaviours on the basis of the models and perceptual input, virtual humans can behave appropriately for a range of social relationships. These kinds of agents have also been referred to by similar terms, including animated agents (Rickel and Johnson, 1999a) or embodied conversational agents (Cassell et al., 2000).

With respect to the dialogue capability, virtual humans have a number of similarities with both task-oriented dialogue systems and chatterbots. Like task-oriented dialogue systems, they generally have knowledge of tasks, and models of the steps involved in the task and how to talk about them. However, generally task-oriented dialogue systems strive to solve the problem as efficiently as possible, minimizing the opportunity for misunderstanding, even if this leads to unnatural and un-human-like dialogue. On the other hand, virtual humans strive for human-like dialogue so as to train communication behaviours that might transfer to real human interaction. Moreover, for training, efficiency in task performance and brevity is not necessarily an advantage – the longer the interaction the more opportunity for learning. Like chatterbots, virtual humans have a focus on believable conversation, but their purpose is not to convince someone that they are actually human, but merely serve as competent role-players to allow people to have a useful interactive experience.

Our virtual humans have been developed incrementally over a number of years, with developments being made in several aspects (Rickel and Johnson, 1999a; Hill, 2000; Rickel et al., 2002; Traum and Rickel, 2002; Traum et al., 2003; Gratch and Marsella, 2004). These virtual humans are embedded in a dynamic virtual world, in which events can happen, agents can perform actions, and humans and virtual humans can speak to each other and communicate using verbal and non-verbal means. The virtual humans are extensions of the Steve agent (Rickel and Johnson, 1999a), and include sophisticated models of emotion reasoning (Gratch and Marsella, 2004), dialogue reasoning (Traum and Rickel, 2002) and a model of team negotiation (Traum et al., 2003). Agents use a rich model of dialogue closely linked with a task model and emotional

appraisals and coping strategies for both interpretation of utterances as well as for decisions about when the agent should speak and what to say.

In previous work (Rickel et al., 2002; Traum et al., 2003), we described a negotiation model that could allow virtual humans to engage as teammates. To negotiate and collaborate with humans and artificial agents, virtual humans must understand not only the task under discussion but also the underlying motivations, beliefs and even emotions of other agents. The virtual human models build on the causal representations developed for decision-theoretic planning and augment them with methods that explicitly model commitments to beliefs and intentions. Plan representations provide a concise representation of the causal relationship between events and states, key for assessing the relevance of events to an agent's goals and for assessing causal attributions. Plan representations also lie at the heart of many reasoning techniques (e.g., planning, explanation, natural language processing) and facilitate their integration. The decision-theoretic concepts of utility and probability are key for modelling non-determinism and for assessing the value of alternative negotiation choices. Explicit representations of intentions and beliefs are critical for negotiation and for assessing blame when negotiations fail (Mao and Gratch, 2004).

This model assumed that teammates shared common end goals, participated in a social institution with roles that the participants played, and had strong trust in the other teammates' abilities and veracity. It did not address how virtual humans might interact in the case where these factors were lacking, and how to begin to form them through interaction.

In this chapter, we extend the dialogue model to allow for non-team negotiation. The extended model allows for the case in which relationships may need to be developed during the interaction, and in which the virtual human's behaviour may be very different depending on the nature and strength of the relationships. We also present Dr Perez, an implemented virtual human who uses this model to negotiate in a prototype training application.

In the next section, we describe the information state dialogue model for virtual humans. This includes both aspects of information state and dialogue moves. In Section 3, we describe how this model is used in understanding and producing communicative behaviour. In Section 4, we discuss non-team negotiation. After a brief survey of literature in the area, we describe our domain testbed and then our first synthesis of this work in terms of strategies for virtual humans, and then extensions to the dialogue model to make use of these strategies. In Section 5, we show two example interactions with this agent, showing how the dynamic trust model is developed during the interaction and how this can affect the agent's choice of utterance. We conclude with some brief remarks about evaluation and future directions.

2. Dialogue Model

Our virtual human dialogue model uses the *Information state approach* (Larsson and Traum, 2000; Traum and Larsson, 2003). In this approach, dialogue is modelled using the following aspects:

- An Information State – including representations of the information used to model dialogue context, distinguishing one (point in a) dialogue from another
- A set of dialogue moves, which represent contributions to dialogue and packages of change to the information state
- A set of rules (or other decision procedures) for modelling the dynamics of dialogue, including the following types of rules:
  - Recognition rules – that interpret raw communication input (e.g., speech, text, gestures) as dialogue moves
  - Update rules – that govern the change in information state based on observation of dialogue acts
  - Selection rules – that choose a set of dialogue acts to perform, given a configuration of the information state
  - Realization rules – that produce communicative output behaviour that will perform the set of dialogue moves

  Rules have a condition part (that specifies constraints on the information state that must be satisfied in order for the rule to fire) and an effect part (that specifies how the information state changes when the rule applies)
- An algorithm that specifies the order and priority of rule application

There are several toolkits that allow one to specify an information state, dialogue moves, rules, and an algorithm, in order to create an information state dialogue system. These include TrindiKit (Larsson et al., 1999), Dipper (Bos et al., 2003) and Midiki (Midiki Users Manual, 2005). Rather than using one of these toolkits, our dialogue manager is implemented in SOAR (Laird et al., 1987). Like these information state toolkits, SOAR has an information state, consisting of objects with links to values and other objects. In this sense it is very much like the information state of Godis (Cooper and Larsson, 1999) and EDIS (Matheson et al., 2000) which are based primarily on AVM-like record structures. SOAR also is a rule-based language. SOAR's main algorithm is to apply all rules simultaneously, and order of application is achieved by referring to dynamic aspects of the information state in the condition parts of the rule. For example, if rule 1 has a condition that requires the presence of a particular

value in the information state and that value is only set by rule 2, then rule 2 will fire before rule 1. While the main body of dialogue processing is achieved by application of rules in SOAR, there are also other computational mechanisms that can be used, e.g., general programs in TCL, and an input/output interface that can send and receive information from external system modules written in any language.

There are two main differences in our virtual human dialogue model that distinguish it from most other information state based dialogue managers. First, the information state and sets of dialogue moves are divided into a number of *layers*, each covering a different aspect of communication (Traum and Rickel, 2002). We believe the scope and breadth of these layers exceeds any other implemented dialogue system in terms of the range of phenomena modelled, allowing our virtual humans to engage in multiparty dialogue, multiple, temporally overlapping conversations, and both team and non-team negotiation. Second, many other parts of the virtual human model, including task reasoning, planning, emotion reasoning, and goal-directed behaviour are also represented in the same information state approach within SOAR as the dialogue model, allowing very rich interaction between these components. Dialogue rules may make use of these aspects of the information state in all phases of processing, from recognition of dialogue moves to generating behaviour.

In the rest of this section, we give an overview of the aspects of information state and dialogue moves that are most important for dialogue processing. In the next section we overview the arrangement of dialogue processing rules.

2.1 Information State Aspects

The top level of the dialogue information state includes a number of aspects including **Ontology, Lexicon, Participants, Social State, Speech Event History, Conversation(s)**, and **Social Planning**. The ontology contains mostly static information about subcategorizations, including selection restrictions of roles for events, and group membership. The lexicon maps words from English and the external recognisers to the internal task and dialogue ontology. The participants list keeps track of all participants (both real and virtual) in the simulation, including information about distance and accessibility for contact, and hypotheses about current gaze and attention of the participants. Social state information includes both the roles and relationships that participants hold to tasks and each other, as well as the obligations and social commitments to propositions that participants hold toward each other.

Multiple conversations can be active at a time, and each one has its own internal structure. Conversation structure includes

- A list of participants in the conversation (who are assumed to understand the grounded contributions), divided into active participants who

perform speaker and/or addressee roles in utterances of the conversation, and overhearers (who do not)

- Modality of the conversation (face to face, radio, etc.)
- The turn-holder (a participant, or none)
- The initiative-holder (a participant or none)
- The purpose of the conversation (e.g. to negotiate a task), if any
- A history of utterances that are part of the conversation
- A history of concept mention
- A structure of questions under discussion
- A grounding structure, consisting of a bounded stack of common ground units (Traum, 1994)

The social planning structure contains information useful for planning and recognising future dialogue actions. The main aspects are:

- A set of *potential obligations*, including actual discourse obligations (Traum and Allen, 1994), as well as those that would be established if an open grounding unit were to be grounded and those that would be established based on conditional rules if the antecedent is planned
- A set of expectations of what is likely to be said next, following from what has been said (e.g., reactions to a suggestion, or discussion of a next step in a plan after the current topic of discussion
- An agenda of partially ordered dialogue goals

The goals on the agenda can come from domain goals in the task model (including various types of communication, such as getting another agent to do something, agreeing on a solution, getting permission, or seeking knowledge), the emotion model, or other aspects of the agent's reasoning process. The agenda is used both for generating new initiatives (see Section 3.1), and as a further source of expectations for use in interpreting utterances that do not refer to the context of what has been recently said or observed, in a manner similar to the account of plan and question accommodation in Larsson (2002).

In addition to these aspects of the dialogue information state, dialogue processing also makes use of a number of information state elements from other modules, including a causal history of past events, the current world state, and plans. Also used are assessments of utility of possible actions and emotional appraisals of potential actions.

2.2 Dialogue Moves

The dialogue model includes multiple layers of interaction, each with associated parts of the information state and dialogue moves. These layers are described in more detail in Traum and Rickel (2002). Figure 1 shows the set of acts in each layer used in the current implementation.

The forward and backwards acts together are classed as *core speech acts*, while the other classes are grouped together as other *dialogue acts*. Core speech acts are most directly connected to the social state part of the information state, adding and relieving obligations, social commitments, and affecting social relations. These acts also have functions related to influencing the topics under discussion in the conversations that they are a part of.

Core speech acts have a content which is either a state, an action description or a question about one of these. Each of the states and actions in the task model is annotated with semantic information that can be used to describe and recognise description of those states in natural language (and our speech-act based agent communication language). Speech recognition and natural language interpretation produces similar contents from spoken utterances. Dialogue processing then compares the NL representation to the relevant task model representations, and, if a sufficiently close match can be found with a task model state or action, that is seen as the referent.

Unlike many accounts of the effects of these speech acts (e.g. Cohen and Perrault, 1979; Allen, 1983; Cohen and Levesque, 1990; Fipa, 1997), there are no direct effects on the beliefs, desires or intentions of the conversational participants. This allows for the possibility that participants are insincere in their utterances. Following Traum and Allen (1994), the direct effects involve social commitments, and one may then infer from these commitments the beliefs or intentions commonly associated with these utterance types, given additional assumptions.

forward acts	assert, info-req, order, request, thank, greeting, closing, express, check, suggest, promise, offer, apology, encourage, accuse, intro-topic, avoid
backward acts	accept, reject, address, answer, divert, counterpropose, hold, clarify-parameter, redirect, confirm
conversation	start-conversation, end-conversation, confirm-start, deny-start, pre-close
grounding	initiate, continue, repair, acknowledge, request-repair, cancel
turn-taking	keep-turn, hold-turn, release-turn, assign-turn
initiative	take-initiative

Figure 1. Types of dialogue moves.

Assertions will have the effect of establishing a commitment by the speaker that the state holds, or that action happened, is happening, will happen, or should happen, depending on the tense and aspect of the utterance. **Info-requests** have a question as their contents. Questions are (possibly partial) propositions together with a designated *q-slot* indicating the part of the proposition asked about. Info-requests have as their effect an obligation to address the question. **Requests** have an action as content, and the effect is an obligation to address the request, e.g., to consider and give feedback on the request. **Orders**, which can only be performed by a superior to a subordinate in the social structure, have as their effect an obligation to perform the action that is its content. **Suggestions** do not impose obligations, but do focus the topic on the action.

3. Dialogue Processing

Language processing occurs in two distinct and interleavable "cycles", one for understanding language and updating the information state, and a second for producing language. This separation of input and output processing cycles allows the agent to have an arbitrary interleaving of contributions by itself and others rather than enforcing a rigid turn-alternation. Each communicative contribution is simultaneously interpreted at each layer, and may correspond to a number of acts at different layers. The interpretation cycle includes stages for speech recognition, semantic parsing, contextual processing (including reference resolution, intention recognition and dialogue act interpretation), and finally updating the information state.

Generation usually starts from an intention to perform one or a small set of acts, however any realized utterance will also correspond to a number of acts, some of which (e.g., turn-taking) may be as much a result of the timing of the performance with respect to other events as to the planned behaviour. Generation proceeds from one of two sources: *reactions* to utterances of others and events in the virtual world, and *initiatives* which proceed from the agents own goals and agenda.

There are different sorts of reactions which are prompted by different aspects of the dialogue information state, including returning greetings, grounding understood material and repairing material that was not understood, addressing obligations, and reacting to proposals. Some are predicated on the agent having the turn (or at least another agent not having the turn), while others (e.g., repairing errors, reacting to danger) are not, and can produce interruptions. There is also a partial order priority scheme so that, e.g., addressing an obligation takes priority over merely acknowledging that a question has been asked.

3.1 Initiative Model

The initiative model consists of three main components, which handle three central problems:

- What to talk about
- When to talk about it
- How to talk about

The first component is modelled by the agenda, mentioned in Section 2.1. Whenever the agent forms communicative goals, these are added to the agenda. Goals on the agenda may be ordered with respect to each other, and special track is kept of all possible next items and the current agenda item that the agent is focused on. For each item, a record is also kept of how many times and in which ways this item has been talked about.

There are multiple ways to trigger the agent to take the initiative. Policies can be set to monitor certain conditions of the dialogue, other mental state, or the environment. These initiative policies can also be individually turned on or off according to a threshold for initiative level, and a current initiative level that is part of the agent's personality profile. Currently the following policies are used:

- When a threshold for too much silence has been exceeded: this is used to insure that a conversation will not stagnate, even if a user does not know what to say.
- When a threshold for too many cumulative errors in understanding has passed: this is used to take control when the user is having problems understanding or making him or herself understood.
- When a threshold for too many consecutive irrelevant utterances has been exceeded: this is used to ensure that the conversation does not drift off the topic of the conversation. In some cases a user may be trying to speak relevantly, but goes beyond the comprehension capability of the agent, either by using unknown vocabulary, constructions, or implicit connections that the agent's inferential power is unable to connect. In all of these cases having the agent take the initiative often makes the conversation more fluent.
- At the directive of a human controller or director agent: this is used to trigger initiative "manually" or for some other reason outside of the reasoning of the agent itself.

When an initiative trigger is reached, the agent will choose an item at the top of the agenda (usually the current item, if there is one), and proceed to take

the initiative. There is still the issue of how to address the agenda item. This depends in part on what kind of item it is and other factors of the domain and context for the conversation. In general, though, there is a cascade of several different modes and a count of how many times an agent will use that mode. For instance, in the MRE domain (Traum et al., 2004), when the Sergeant wants to propose an item that he would like the Lieutenant (his superior, but also his trainee) to do, he sequentially uses the following modes to bring up the desired action:

1 Hint: mention an end goal of the action (according to the task model) that is currently unfulfilled, or a pre-condition of the action that has already been met, thus enabling the action.

2 Suggest: directly suggest the action itself as a possibility.

3 Request: specifically request permission to perform the action.

4 Perform: perform the action without authorization (unless specifically prohibited).

Together, these factors of the initiative model allow the agents to engage in mixed-initiative dialogue, with the level of initiative that the agent takes being a factor both of customizable parameters as well as dynamic conditions of the dialogue.

4. Non-Team Negotiation

The model presented in the previous sections was designed mainly for team interaction, where it is assumed that the teammates have the same general goals, although they may disagree and negotiate about the best ways to achieve those goals. For more general situations, we must generalize the model of negotiation to include neutral and adversarial conditions.

4.1 Orientations Toward Negotiation

One of the central ways to characterize negotiation under adversarial conditions is with respect to the tension between competition and cooperation. Negotiators may have different goals, perceive themselves in conflict over those goals but may also perceive the need to cooperate to some degree to achieve their goals. In this view, one can characterize the state of a negotiation process from the perspective of the competitive/cooperative orientation of the parties to the negotiation and the strategies they employ in light of those orientations. Specifically, one oft-made distinction is between integrative and distributive (Walton and Mckersie, 1965) situations. If a negotiation is a win–lose game where there is a fixed value to be distributed, then it is called distributive.

There will be a winner and a loser. In contrast, an integrative situation is one where both sides can potentially win, a win–win situation where negotiation could add value and be of benefit to both sides. These basic distinctions presume some commitment to engage in negotiation. However, an individual may simply believe that there is no possible benefit or even need to negotiate. This individual may have an orientation to simply avoid the negotiation or deny the need for it, what is termed avoidance (e.g. Sillars et al., 1982). We thus start with three basic orientations toward a negotiation: avoidance, distributive, and integrative. Whenever an agent seriously considers a negotiation situation it will choose one of these three orientations.

Negotiators may perceive a situation as one to be avoided, or as a distributive or integrative situation regardless of whether this reflects the true situation. Changing the perceptions of other agents is often one of the main tasks in a successful negotiation. Based on current perceptions, people tend to use a range of dialogue tactics consistent with their orientations (Putnam and Jones, 1982; Sillars et al., 1982). Avoidance tactics include shifting the focus of conversation and delays. Distributive tactics can include various defensive moves such as stating prior commitments that bind the negotiator or arguments that support the negotiator's position. Distributive tactics can also be more offensive, such as threats, criticisms and insults. Integrative tactics are more cooperative with negotiators actually attempting to see issues from the other's perspective. Tactics can be arguments that support the other's position, acceptances of offers, offers of support, etc. Note at a finer grain of analysis, the tactics employed have both instrumental and affective components. For example, distributive tactics, besides trying to gain competitive advantage, tend to be associated with angry or intimidating behaviour whereas the integrative tactics try to promote a positive affective climate (Putnam and Jones, 1982).

Negotiators will often shift orientations during the course of a negotiation. Several factors have been identified as being critical to moving towards an integrative orientation, including acts of reciprocity, establishing trust and reinforcing shared goals (e.g. Wilson and Putnam, 1990).

4.2 Domain Testbed: Support Operations

Whether it is Kosovo, East Timor, or Iraq, one lesson that has emerged from attempts at "peacemaking" is that negotiation skills are needed across all levels of civilian and government organisations involved. To have a lasting positive effect, interactions between military and locals must be carried out in a way that generates goodwill and trust. We have selected this general class of operations as a testbed for our work on negotiation.

More specifically, we are developing a training scenario in which a local military commander (who has a rank of captain) must negotiate with a medical

Figure 2. VR clinic and virtual human doctor.

relief organisation. A virtual human plays the role of a doctor running a clinic. A human trainee plays the role of the captain, and is supposed to negotiate with the doctor to get him to move the clinic, which could be damaged by a planned military operation. Ideally, the captain will convince the doctor without resorting to force or threats and without revealing information about the planned operation. Figure 2 shows the trainee's view of the doctor in his office inside the clinic. The success of the negotiation will depend on the trainee's ability to follow good negotiating techniques, when confronted with different types of behaviour from the virtual doctor.

4.3 Negotiation Strategies for Virtual Humans

One of our first steps toward implementing a virtual doctor character was to analyze how people act in that role. To this end, we have conducted a series of role-play sessions, in which one person plays the role of the captain while another plays the role of doctor. Each is given a short set of instructions with different background information, goals, and resources for the negotiation, but given freedom as to how to conduct the negotiation and react to their partner. In these dialogues we can see examples of each of the orientations described in the previous section. For example, in the first one, the doctor displays an avoidance orientation, and is able to divert the topic of the conversation from the move to the military's role in upcoming operations for over 10 turns (only the first few are shown here). In the second one, we see a doctor illustrating the

distributive orientation, contesting the basic facts and goals rather than working together on common issues. In the third one, we see an example of integrative orientation, the doctor having accepted the danger of the current location and willing to meet the captain's goals if his own are also addressed.

(1) C: It's a temporary move, once the battle is over, you will be moved back.
D: Why don't you cancel your battle? Why don't you not kill these people.
C: We're not the ones deciding the battle.
D: You're the ones here. You're telling me this.

(2) C: We need to move as soon as possible. There are insurgents in the area. This is very unsafe, you're putting yourself and your patients in danger.
D: Why? I don't want to move. I have all these patients here. They won't move, if I move who would who could save them?
C: Sir, everyone is in danger! If we stay here there's ...
D: I'm not in danger.

(3) C: Insurgents will not hesitate to harm civilians if that's their path that they need to take. They won't hesitate to harm doctors, a doctor or even injured patients if they feel that's the the means to their end.
D: Well.
C: This is why you need to come to us.
D: I think we can make a deal. You can give me medical supply, and then we can go with you. I need supplies as soon as possible. As you can see, we are running out of supplies.

We have developed *strategies* for each of these orientations. Our virtual humans can use the strategies to adjust their behaviour toward the orientations described above. A strategy consists of several aspects including: **entry conditions**, which indicate when adoption is appropriate; **exit conditions**, which indicate when the strategy should be dropped (often in favour of more appropriate strategies); **associated moves**, which can be performed as tactics to implement the strategy; and **influences** of the strategy on behaviour and reasoning. These aspects result from the underlying emotion and dialogue models of the virtual humans.

The EMA (**EM**otion and **A**daptation) model of emotion (Gratch and Marsella, 2004) describes how coping strategies arise as cognitive and physical responses to important events, based on the appraisal (Scherer et al., 2001) of perceptions related to goals and beliefs. Appraisal characterizes events in terms of variables that guide the selection of an appropriate response (e.g., is this desirable? can it be avoided?), but the event need not be physical. Negotiation

strategies can thus be seen as types of coping strategies in which the event in question is the negotiation itself, and moves are the types of dialogue actions an agent will perform as part of a negotiation.

The avoidance orientation arises from an appraisal that the negotiation is undesirable but avoidable. The main motivation is to try to escape from the negotiation. When this appraisal is active, the agent chooses an **avoidance** strategy. Exit conditions will be the negation of either of the entry conditions — when the agent believes either that the negotiation has some utility or that it is not avoidable, the agent will abandon the avoidance strategy. The avoidance strategy involves attempts to change the topic of a conversation or get out of it entirely. When applying the avoidance strategy an agent will refrain from commenting on the object of negotiation, even to refute claims.

When in distributive mode, the agent will attempt to "win" rather than "lose" the negotiation. This can be associated with several strategies, depending on the type of decisions to be made and the range of possible alternatives. An *attack* strategy is appropriate when the appraisal is that a negotiation is not avoidable and the proposal is undesirable. Other strategies are also appropriate for a distributive orientation, including defence against a threat rather than attack, or making unreasonable demands in the hope the other party will drop the negotiation. We defer this for future work. One should drop an attack strategy when either the negotiation becomes desirable, or it becomes more profitable to avoid (or defend) than attack. The attack strategy involves pointing out the reasons why a proposal is flawed, or ad hominem attacks on the negotiator.

An integrative orientation leads to attempts to satisfy the goals of each of the participants. The **negotiate** strategy is appropriate when an agent thinks there is a possible value to the negotiation — e.g., there is a higher expected utility from the expected outcomes than would be the case without the negotiation. This strategy is dropped either when the perceived utility of continuing to negotiate drops below a threshold, or when the negotiation has been completed. Moves in the negotiation strategy involve problem solving and bargaining, much in the manner of the team negotiation in Traum et al. (2003).

The success of a negotiation is also mediated by factors that influence the perceived trust between parties, including a belief in shared goals, credibility and interdependence. The doctor is unlikely to be swayed by an offer of aid if he does not believe the captain can and will fulfil his commitments. Trust issues are pervasive throughout the negotiation, since there is usually not much point in negotiating with someone you expect to lie, be ill-disposed toward you, or not keep their side of a bargain.

4.4 Extensions to the Dialogue Model

Several extensions to the dialogue model were needed to handle possibly adversarial negotiation and the types of phenomena occurring in this domain. The most basic is sensitivity to the dialogue strategy, which involves overriding some basic reaction rules in some cases. For example, when applying the avoidance strategy one must not directly address a proposal that is on the topic of avoidance. Sometimes these utterances are not even grounded as a way of avoiding talking about an unpleasant topic. In this section we will examine two other extensions: updates to the initiative model and a new type of backward dialogue act for intentionally flouting the Gricean maxim of cooperativity (Grice, 1975).

4.4.1 Modelling trust. According to the dialogue model in Matheson et al. (2000), the direct effect of an assertion is the introduction of a commitment, whether or not either party believes in the assertion. While this is sufficient for reasoning about the claims and responsibility for information, we need to go further and potentially change beliefs and intentions based on communicated information. Trust is used to decide whether to adopt a new belief based on the commitments of another.

Similar to Marsella et al. (2004) and Cassell and Bickmore (2003), trust is modelled as function of underlying variables that are easily derived from our task and dialogue representations. *Solidarity* is a measure of the extent to which parties have shared goals. It is derived from a running tally of how many times the trainee makes assertions or demands that are congruent with the agent's goals. *Credibility* is a measure of the extent to which a party makes believable claims. It is derived from a running tally of how many times the trainee makes assertions that are consistent with the agent's beliefs. Finally, *familiarity* is a measure of the extent to which a party obeys norms of politeness. Currently, an overall measure of trust is derived as a linear combination of these three factors.

4.4.2 Extended initiative model. For properly modelling negotiation strategies the initiative model must be changed from that of team collaboration. The basic mechanism remains in place, however the current strategy is also made part of the agenda. Some agenda items are tied to particular strategies. For the avoidance strategy, initiative will concern the agent's own goals that are unrelated to the topic of negotiation. The irrelevance threshold is also disabled, as this strategy "succeeds" when the agent is able to shift the topic somewhere else.

For the attack strategy, each problem that the agent foresees with the action that is the topic of negotiation is added to the agenda. These problems include

- Pre-conditions of the action that are not met
- Other, better plans (according to the utility calculations in the task model)
- Undesirable side effects of the action

Another general agenda item for this domain in particular, is a desire for Doctor Perez to get back to his patients and end the conversation. This generally takes lower priority than more specific items within a strategy, but can come out if no other agenda items are available. Following the model in Section 3.1, there are several levels of utterances to support this goal, ranging from pre-closing reminders that he is busy, to finally ending the conversation.

4.4.3 Avoidance moves. Another important extension to the dialogue model is the addition of an *avoidance* dialogue act. According to Traum and Allen (1994), when presented with a request or question, one has an obligation to address this utterance, though not necessarily accept it and perform the desired act. There are many cases in which one might not want to perform the act. Rejection is an option which addresses the speech act, but may have negative consequences of its own. First, it commits the speaker to a negative position, which may not be desired. Second, it serves as a face threat (Brown and Levinson, 1978). A third option is to try to avoid an explicit commitment. One way to do this is by deferring a resolution to the future (e.g., "let me get back to you about that"), which does also commit one, and also may not be satisfactory to the interlocutor. Another type of action is to attempt to change the topic altogether. Depending on how explicit the request is, this may or may not go against an obligation to address. Thus, mentioning a topic indirectly is more polite than a direct request, because it does not force someone to not meet an obligation when avoiding.

We can see examples of avoidance from both participants in the dialogue in Figure 4. In turns 8, 10, and 12 the doctor is avoiding the move-clinic topic, following an explicit strategy to talk about something else. In turn 11, the captain is avoiding giving a direct answer to the doctor's question about whether the captain will attack. The captain is in a tough situation here: if he says "yes", he will hurt both solidarity (by showing conflicting goals) as well as possibly endanger his mission by revealing future plans. On the other hand, if he says "no", he will be lying which will hurt credibility if he is found out. By avoiding the question, he does not fall into either of these traps, though there is a danger of not fulfilling the discourse obligations. The explanation ("I can not disclose more information") does partially address this issue, and leaves the door open for future trust-building, such as in turn 13.

5. Example Interactions

Consider the dialogue in Figure 3. This is just one of many possible interactions, depending on the choices of the human captain, as well as several aspects (some probabilistic) influencing the choice of moves and strategy transitions of the virtual human doctor. Shown along with the dialogue are strategy transitions and the major influences on the trust components: familiarity, solidarity and credibility.

Here the captain acts as he might with a team member — after starting the conversation, launching directly into his purpose, and answering a question straightforwardly. While this would have worked with a subordinate teammember, it has disastrous effects on the neutral doctor, bringing his trust level almost down to zero and failing to accomplish both objectives.

In this dialogue, nothing was done by the captain to try to establish a better relationship with the doctor, or address the issue of differing objectives and

1	C	Hello Doctor Perez.
2	D	Hello.
3.1	C	I have orders to move this clinic to another location.
4	D	You want to move the clinic?
5	C	Yes
-SOLIDARITY'		
Strategy shift: Avoid negotiating about move		
6.1	D	Look at these people!
6.2	D	We need to help them.
7	C	It is not safe here, we can't protect you.
-CREDIBILITY'		
8.1	D	Protect me? Protect me from what?
8.2	D	Are you going to attack?
9	C	Yes
-SOLIDARITY'		
10.1	D	We need proper supplies here
Strategy shift: Disrupt negotiation about move		
10.2	D	To move the clinic is not possible, we have many patients in critical condition.
11	C	It is very important to move now!
-SOLIDARITY'		
12.1	D	You are the threat, I need protection from you!
Strategy shift: Unsuccessfully conclude the negotiation		
12.2	D	I would have to refuse this decision.
12.3	D	I must go now.

Figure 3. Unsuccessful negotiation dialogue between C, a captain (human trainee) and D, a doctor (virtual human) showing positive and negative effects on trust.

beliefs. The first exchange after the greetings (utterances 2–5) lowers solidarity by showing different objectives, setting up more of an antagonistic than cooperative interaction. The doctor tries to avoid the topic, focusing instead on his patients, rather than the captain's stated goal. The captain tries to argue for his proposed course of action, but only makes things worse with utterance 7. First, he says something the doctor doesn't believe (that the clinic is in danger), lowering his credibility. The doctor is able to reason though that perhaps the captain knows of a reason why it will be unsafe, and challenges by asking if he is going to cause the danger. In 9, the captain answers sincerely, which is a mistake on two fronts. First, he reveals more about his mission than he should to an outsider, possibly endangering its success if word gets out to his enemies. Second, he shows even further divergence from the doctor's goals — attacking rather than helping the patients. After one more brief attempt to change the topic and get help for his own goals, the doctor gives up on the captain in (10.2), and tries to get out of the negotiation. The captain has failed in his objective and prospects are not good for future relations.

For really learning about negotiation it is very helpful to know not just what the other party did, but why. In real negotiations it is usually not possible to get "inside the head" of the negotiating partner, and even subsequent questions can sometimes damage the nature of the interaction itself. In this respect, virtual humans present a real opportunity to improve on training. As shown in Figure 3, we have implemented a trace facility that provides an annotated transcript of the dialogue, showing not just what the virtual human thought was said, but how it influenced his trust, beliefs, and strategy choice. This tool can be used in an "after action review" (AAR) to look in detail at the specific effects the trainee's negotiation tactics had. Here we can see the reasons for decreases in credibility and solidarity as effects of the commitments the captain makes in relation to desires and beliefs of the doctor.

Figure 4 shows a more successful interaction, where the captain tries to build bonds as well as accomplish his task. While the captain's behaviour in this dialogue is not perfect either (the captain might have been better served spending more time up front establishing familiarity and solidarity and perhaps addressing the doctor's concerns first), it is a big improvement over the dialogue in Figure 3. Here the greetings in turn 3 add some familiarity, and the evasion in turn 11 does not do as much damage as the blanket statement of acting against the doctor's interest in the previous dialogue. Things are still not going very well, though, until the captain establishes some common goals with turn 13. With slightly higher trust, the doctor does not break off negotiation at this point, but rather raises a series of objections. By addressing each of the doctor's concerns: safety of patients, lack of supplies, lack of transport, and neutrality, the captain is able to bring him around to the point where the move is not an absolute negative, but is worthy of consideration, as part of a team

1 C Hello Doctor Perez.
2 D Hello.
3 C Very nice to meet you! [+*FAMILIARITY*]
4 D What do you want?
5 C I have orders to assist you in moving this clinic to a safer location.
6 D You want to move the clinic?
7 C Yes [-*SOLIDARITY*]

Strategy shift: Avoid negotiating about move

8.1 D This conflict is madness, it is killing people!
8.2 D We need proper supplies here!
9 C It is not safe here we can't protect you. [-*CREDIBILITY*]
10.1 D Protect me? Protect me from what?
10.2 D Are you going to attack?
11 C I can not disclose more information but it will be a problem to stay here.
12 D You are the problem your bombs are killing these people!
13 C Sir, we are trying to help these people too. [-*CONTROL* +*SOLIDARITY*]

Strategy shift: Disrupt negotiation about move

14.1 D The patients are here,
14.2 D to move the clinic is not possible we have many patients in critical condition.
15 C We can move you to a safer location.
16 D I can't move all these patients.
17 C We can organise the transportation and assist you with the moving. [+*SOLIDARITY*]
18 D I can't move all these patients.
19 C We can have locals move you. [+*SOLIDARITY*]
20 D There are no supplies there.
21 C We can provide you with the supplies you need to help your patients. [+*SOLIDARITY*]

Strategy shift: Willingly negotiate about move

22.1 D I think I understand what you say.
22.2 D Well perhaps we can reach an agreement,
22.3 D but before we can think about moving, we need you to bring antibiotics and plasma to stabilize the patients. [+*INTERDEPENDENCE*]
23 C We can do that! [+*SOLIDARITY*]
24.1 D Well,...
24.2 D Very well captain contact my assistant to make further arrangements.
25 C I'll see to it personally. [+*SOLIDARITY*]
26.1 D I understand your position.
26.2 D My patients need my attention now.
27 C Thank you Doctor!
28.1 D Well,....
28.2 D I must go now
29 C Goodbye.
30 D Good bye.

Figure 4. More successful negotiation.

plan. Finally, the two participants reach an agreement including giving needed supplies as part of the conditions of moving the clinic.

We can see several distinct phases of the dialogue in Figure 4, relating to different negotiation strategies. The initial segment (turns 1–7) includes initial greetings and establishing the topic for the conversation — the captain wants

Figure 5. More relaxed and open doctor.

to move the clinic. In turns 8–12, the doctor engages in an *avoidance* strategy, trying to avoid this topic by bringing up other issues, such as his need for supplies, and the general problems of conflict. In turns $14-20$, the doctor has adopted an *attack* strategy, and points out problems with the proposed move. In turns 22–25, the doctor adopts a more open negotiation strategy, and an actual bargain is struck. Finally, turns 26–30 show a closing phase in which the doctor disengages from the conversation, while the captain tries to establish good relations for future interaction. Application of these strategies influences not just the choice of dialogue move, but the whole body posture of the doctor and use of gestures and expressions as well. For example, when the doctor is feeling more distant and less trusting, he adopts a closed posture (Figure 2). When he is more trusting and open to negotiation, the posture becomes more relaxed (Figure 5).

6. Preliminary Evaluation and Future Directions

As part of the development of the system, we have so far tested the system with over 50 different people acting as a trainee, most with several dialogues. As of this writing we have not fully analyzed this data, however we can draw some general conclusions. Usually people are able to have a coherent dialogue with the Doctor, although some problems arise when concepts are brought up that are beyond his vocabulary. An advantage of this domain is that when the doctor is following an avoidance or attack strategy, it is natural for him to take the initiative and complain about his own concerns rather than being directly

responsive to the human's utterance, so some non-understandings do not lead to problems. Most people who talk to the doctor do not convince him in the first session, however, a little bit of explaining of proper negotiating techniques (e.g., build trust before arguing) generally leads to a successful negotiation in a follow-up round.

Future work involves extension of the models to include additional negotiation strategies, emotion-based styles of interaction within the strategies, and application to other scenarios, some involving cultural differences in behaviour and interpretation, as well as translated and multilateral dialogue.

Acknowledgements We would like to thank the entire SASO-ST team for producing the system in which this Dialogue and Negotiation model has been embedded. The work described in this chapter was supported by the Department of the Army under contract number DAAD 19-99-D-0046. Any opinions, findings and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the Department of the Army.

References

Allen, J. F. (1983). Recognizing Intentions from Natural Language Utterances. In Brady, M. and Berwick, R. C., editors, *Computational Models of Discourse*, pages 107–166. MIT Press, Cambridge.

Bos, J., Klein, E., Lemon, O., and Oka, T. (2003). Dipper: Description and Formalisation of an Information-State Update Dialogue System Architecture. In *Proceedings of SIGdial Workshop on Discourse and Dialogue*, pages 115–124, Sapporo.

Brown, P. and Levinson, S. (1978). Universals in Language Usage: Politeness Phenomena. In Goody, E. N., editor, *Questions and Politeness: Strategies in Social Interaction*, pages 56–311. Cambridge University Press, Cambridge.

Cassell, J. and Bickmore, T. (2003). A Relational Agent: A Model and Implementation of Building User Trust. In *Proceedings of ACM CHI Conference*, pages 396–403, New York.

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

Cohen, P. R. and Levesque, H. J. (1990). Rational Interaction as the Basis for Communication. In Cohen, P. R., Morgan, J., and Pollack, M. E., editors, *Intentions in Communication*. MIT Press, Cambridge.

Cohen, P. R. and Perrault, C. R. (1979). Elements of a Plan-Based Theory of Speech Acts. *Cognitive Science*, 3:177–212.

Cooper, R. and Larsson, S. (1999). Dialogue Moves and Information States. In *Proceedings of International Workshop on Computational Semantics (IWCS)*, pages 398–400, Tilburg.

Fipa (1997). Fipa 97 Specification Part 2: Agent Communication Language. http://drogo.cselt.stet.it/fipa/spec/fipa97/f8a21.zip.

Gratch, J. and Marsella, S. (2004). A Domain-Independent Framework for Modeling Emotion. *Journal of Cognitive Systems Research*, 5:269–306.

Grice, H. P. (1975). Logic and Conversation. Syntax and Semantics. In Cole, P. and Morgan, J. L., editors, *Speech Acts*, volume 3, pages 41–58. Academic, New York.

Hill, R. (2000). Perceptual Attention in Virtual Humans: Towards Realistic and Believable Gaze Behaviors. In *Proceedings of AAAI Fall Symposium on Simulating Human Agents*, pages 46–52, North Falmouth.

Laird, J. E., Newell, A., and Rosenbloom, P. S. (1987). SOAR: An Architecture for General Intelligence. *Artificial Intelligence*, 33:1–64.

Larsson, S. (2002). *Issue-Based Dialogue Management*. PhD thesis, Göteborg University, Sweden.

Larsson, S., Bohlin, P., Bos, J., and Traum, D. (1999). *Trindikit Manual*.

Larsson, S. and Traum, D. (2000). Information State and Dialogue Management in the TRINDI Dialogue Move Engine Toolkit. *Natural Language Engineering,Special Issue on Spoken Language Dialogue System Engineering*, 6:323–340.

Mao, W. and Gratch, J. (2004). Social Judgment in Multiagent Interactions. In *Proceedings of 3rd International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 210–217, Columbia University, New York.

Marsella, S., Pynadath, D., and Read, S. (2004). Psychsim: Agent-Based Modeling of Social Interactions and Influence. In *Proceedings of International Conference on Cognitive Modelling*, pages 243–248, Pittsburgh.

Matheson, C., Poesio, M., and Traum, D. (2000). Modelling Grounding and Discourse Obligations using Update Rules. In *Proceedings of Conference of the North American chapter of the Association for Computational Linguistics (NAACL)*, pages 1–8, Seattle.

Midiki Users Manual (2005). Mitre Corporation. https://sourceforge.net/docman/display_doc.php?docid=25561&group_id=123104.

Putnam, L. L. and Jones, T. S. (1982). Reciprocity in Negotiations: An Analysis of Bargaining Interaction. *Communications Monograph*, 49:171–191.

Rickel, J. and Johnson, W. L. (1999a). Animated Agents for Procedural Training in Virtual Reality: Perception, Cognition, and Motor Control. *Applied Artificial Intelligence*, 13:343–382.

Rickel, J. and Johnson, W. L. (1999b). Virtual Humans for Team Training in Virtual Reality. In *Proceedings of 9th International Conference on Artificial Intelligence in Education*, pages 578–585, Le Mans.

Rickel, J., Marsella, S., Gratch, J., Hill, R., Traum, D., and Swartout, W. (2002). Toward a New Generation of Virtual Humans for Interactive Experiences. *IEEE Intelligent Systems*, 17:32–38.

Scherer, K. R., Schorr, A., and Jonstone, T. (2001). *Appraisal Processes in Emotion*. Series in Affective Science. Oxford University Press, Oxford.

Sillars, A. L., Coletti, S. F., Parry, D., and Rogers, M. A. (1982). Coding Verbal Conflict Tactics: Nonverbal and Perceptual Correlates of the Avoidance-Distributive-Integrative Distinction. *Human Communication Research*, 9(1):83–95.

Traum, D. R. (1994). *A Computational Theory of Grounding in Natural Language Conversation*. PhD thesis, Department of Computer Science, University of Rochester, USA. Also available as TR 545, Department of Computer Science, University of Rochester, New York.

Traum, D. R. and Allen, J. F. (1994). Discourse Obligations in Dialogue Processing. In *Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 1–8, Las Cruces, New Mexico.

Traum, D. R. and Larsson, S. (2003). The Information State Approach to Dialogue Management. In van Kuppevelt, J. and Smith, R., editors, *Current and New Directions in Discourse and Dialogue*, pages 325–353. Kluwer Academic.

Traum, D. R. and Rickel, J. (2002). Embodied Agents for Multi-Party Dialogue in Immersive Virtual Worlds. In *Proceedings of 1st International Joint conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 766–773, Bologna.

Traum, D. R., Rickel, J., Marsella, S., and Gratch, J. (2003). Negotiation over Tasks in Hybrid Human-Agent Teams for Simulation-Based Training. In *Proceedings of Second International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 441–448, Melbourne.

Traum, D. R., Robinson, S., and Stephan, J. (2004). Evaluation of Multi-Party Virtual Reality Dialogue Interaction. In *Proceedings of International Conference on Language Resources and Evaluation (LREC)*, pages 1699–1702, Lisbon.

Walton, R. E. and Mckersie, R. B. (1965). *A Behavioral Theory of Labor Negotiations: An Analysis of a Social Interaction System*. McGraw-Hill, New York.

Wilson, S. R. and Putnam, L. L. (1990). Interaction Goals in Negotiation. *Communication Yearbook*, 13:374–406.