Applied Artificial Intelligence, 19:287–311

ISSN: 0883-9514 print/1087-6545 online
DOI: 10.1080/08839510590910183

AFFECT IN TUTORING DIALOGUES

Dirk Heylen, Anton Nijholt, and Rieks op den Akker □ *Department of Computer Science, University of Twente, Enschede, The Netherlands*

□ *This paper is about INES, an intelligent, multimodal tutoring environment, and how we build a tutor agent in the environment that tries to be sensitive to the mental state of the student that interacts with it. The environment was primarily designed to help students practice nursing tasks. For example, one of the implemented tasks is to give a virtual patient a subcutaneous injection. The students can interact multimodally using speech and a haptic device under the guidance of the virtual embodied tutor. INES takes into account elements of the student's character and an appraisal of the student's actions to estimate the mental state of the student. This information is used to plan and execute the actions and responses of the tutor agent.*

Successful social interaction requires that every interactant has a good sense of what the other believes and feels. The way we interact and the way we feel is heavily influenced by what we believe about the beliefs, intentions, and emotions of our partners in interaction and how we want to influence these. This concerns both attitudes, in general, about the topic of conversation, and attitudes towards the interactant in particular. This determines how we present ourselves; what we reveal, hide, and fake. Also, facial expressions and other forms of nonverbal communication, depend mostly on social factors (Heylen 2003). In short, the theories we have about another are crucial to the way we operate in interactions. The same should be the case for artificial virtual humans that are designed for interaction.

When building embodied conversational agents that can hold their part of the conversation, we want not just to experience them as task-oriented dialogue systems with a face added on, but as social actors playing more complex interaction games. We want to design them taking a broader perspective, incorporating what we know about behavior during social interaction, accounting for the various factors that play a role in interpersonal relations. We would therefore like to see agents with modules that

Address correspondence to Anton Nijholt, University of Twente, Department of Computer Science, P.O. Box 217, 7500 AE Enschede, The Netherlands. E-mail: anijholt@cs.utwente.nl

are specifically concerned with reading the minds of others as a form of "social intelligence" (Baron-Cohen 1995). In many cases, this means that we should pay special attention not just to what the characters themselves should be feeling, but more importantly about how they think the other feels, part of which is reasoning about such things as "how will the other feel if I would show (or hide) what I'm thinking and feeling and what would I feel about this effect." Rather than dealing with raw affect, emotional intelligence involves knowing how to play an interpersonal, social game and means being sensitive to the feelings of others (Goleman 1997). With this knowledge, the agents should be able to deliberately choose a tactic of what to say and show (including, for instance, emotions felt or faked).

Although we may be far removed from agents that can read the minds of the persons they interact with, we can already start to design them in such a way that they take into account what they believe to be the mental state of the other when planning and executing their actions and responses, even though these beliefs may not be based on the same mind-reading capabilities of humans.

In our work on INES, an intelligent, multimodal tutoring environment, we have started to build a tutor agent that tries to be sensitive to the mental state of the student that interacts with it. Tutoring situations are essentially a social encounter, the goal of which is for a student to learn some task or acquire knowledge with the tutor acting in all kinds of ways to assist the student with this goal. The actions of a tutor are also not just restricted to pure instructions, but they should also create the right emotional conditions for a student to act. The fact that the tutoring situation is a social encounter means that influencing the emotional state proceeds through social acts with emotion-changing potential. For instance, the tutor has the status to judge (criticize or praise) the student for his actions.

INES is an intelligent tutoring system that was primarily designed to help students practice nursing tasks using a haptic device within a virtual environment. We have paid special attention to affective control in the tutoring dialogues by selecting the appropriate feedback. The kind of action, the affective language used, and the overall teaching tactics are adjusted. INES does not have extended mind-reading capabilities or elaborate sensor systems for computing the affective state of the student. Instead INES takes into account elements of the student's character, his or her confidence level, for instance, and an appraisal of the student's actions: Did the student make many mistakes, how harmful are the errors made, how was the overall performance so far, etc., to estimate the mental state of the student. It also takes into account the emotional effects the feedback the system itself provides.

The work described in this paper is an extension of our earlier work on tutoring systems, particularly the INES system (Hospers et al. 2003) and the work we did on emotion modeling as described in Bui et al. (2002); Burghouts et al. (2003); and Vissers (2003). In the generic INES system, the "mind" of the tutor was set up as a society of agents representing different cognitive functions or specialized "expert" knowledge. In this paper we shall describe how this system was extended with components that take care of the tutor's social and emotional intelligence skills.

THE INES SYSTEM

The Intelligent Nursing Education Environment System (INES) is an application that allows students to practice nursing tasks. An example of such a procedural task is to give a virtual patient a subcutaneous injection. This was implemented as our first exercise. Such a task requires the execution of several subtasks, for example, taking care that the instruments are sterilized, that there is communication with the patient, and that the injection is done in a correct way. The students can interact multimodally using speech and haptic device,[1] amongst other modalities, under the guidance of a virtual embodied tutor.

INES has been built using our own multiagent platform. The virtual tutor receives input from different agents, for example, from error agents that keep track of what the student is doing with the haptic device. The input for the virtual tutor in the current INES environment consists of keystrokes, mouse movements and clicks, movements, and force using the haptic device and speech. The multimodal input, embedded in its situational and dynamic context, allows the tutor agent to make assumptions and from that compute possibly emerging emotions of the student that is performing a (sub-)task in the INES environment.[2] The virtual tutor can be represented as an embodied agent in the form of a talking face. The tutor can display its approval or disapproval by its facial expressions. It interacts with the student using speech recognition and speech synthesis. The student can also communicate with the patient. For example, for this particular application, asking her to move her arm or to roll up a sleeve. Figure 1 shows the current INES configuration: a student interacting with the system using speech, keyboard and haptic device, the virtual tutor, and the virtual patient.

The architecture of this educational system is generic both with respect to the teaching tasks (the exercises) that can be implemented and the interface options. The system allows for several interaction modalities to connect the student to the different components of the system. The configuration shown in Figure 1 is just one of the many that are possible. The general INES architecture is presented in Figure 2. The architecture is based on two architectural styles:

FIGURE 1 Multimodal tutoring environment INES.

- Data abstraction and object-oriented organization: Each component in INES is represented by an object that has its own attributes and methods.
- Event based, implicit invocation: Agents are implemented and each agent is capable of transmitting (or receiving) messages to other agents.

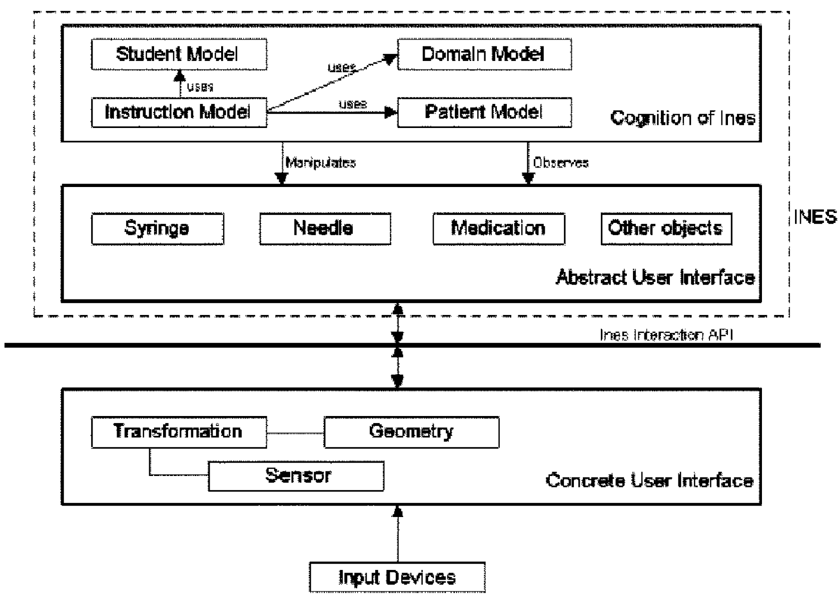

FIGURE 2 INES system architecture.

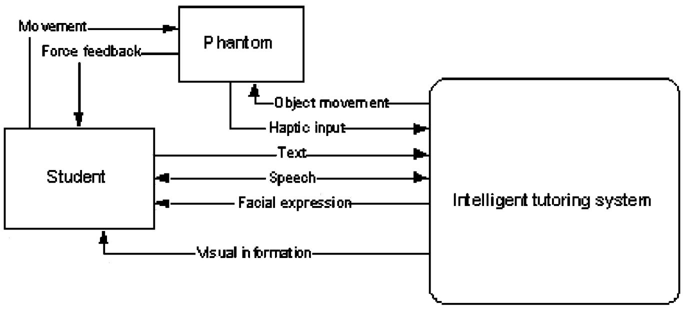

FIGURE 3 Context diagram of INES.

The *concrete user interface* takes care of the visualization of the exercises for the user. The *abstract user interface* is a high-level model of the concrete user interface (e.g., keeping object data like positions). In Figure 2, the *cognition of INES* represents the cognitive knowledge and skills of the system.

Students interact with the system through an interface that makes information available about the patient, the current task and its subtasks. Part of the screen contains a view on the task in progress. The system was designed to support different interfaces. In Figure 3, we display a context diagram of our system.

In a first implementation of the exercise, the system showed video clips. The current version uses the interface modules outlined in Figure 1 for multimodal communication. The haptic device is used to perform a variety of actions: Wiping a piece of cotton along the arm for disinfections or to manipulate the needle when giving the injection. In Table 1, we give an overview of the tasks that have to be performed by the student.

Certain actions cannot be performed by the haptic device. For instance, one of the subtasks that has to be performed involves correctly positioning the patient and her arm in order to be able to give the injection. The student can use speech input to perform this subtask. Speech recognition has been implemented using SpeechPearl and an agent has been introduced in the platform that takes care of the communication with the recognizer. The recognizer employs a grammar in which possible utterances have been specified. Examples of utterances are:

- Put your arm on the rest.
- Would you please put your left arm on the rest?
- Please, bare your right upper arm.
- Roll up your right sleeve.

These interactions are between student and patient. The patient is shown in Figure 4.

TABLE 1 Subtasks with Their Conditions and Checks

Subtask	Conditions indicating the student is working on this specific subtask	Checks needed
Positioning of the virtual patient through speech.	Speech detection.	Check if the correct utterances are said by the student.
Disinfecting the injection area.	Selection of the sterile cotton contained within a pair of tweezers, which is used as avatar for the PHANToM and as disinfecting equipment.	Check if the virtual patient was positioned correctly. Check if the correct area is being disinfected.
Actual injecting of the needle in the injection area.	Selection of the syringe, which is used as avatar for the PHANToM and for the injection.	Check the injection angle and speed. Check if the needle touches only the correct upper arm, which needs to be injected into. Check whether the injection is done in an area that is disinfected.
Administering of the medication through the injected syringe.	Pressing the button available on the PHANToM.	Check whether the needle is within the injection area. Check whether the injection depth is correct (skin layer).
Removal of the needle from the injection area.	The student has just pressed the button on the PHANToM and medication is administered.	Check if there has elapsed a ten second waiting period before the needle is retracted.

There are also interactions between student and tutor. Depending on the progress and the errors made by the student, the tutor chooses dialogue acts, that is, responses to previous interaction activity by the student.

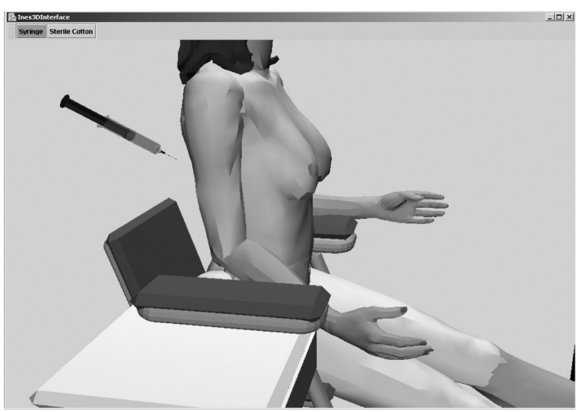

FIGURE 4 INES 3D interface: patient.

These responses are chosen from available hints ("Take care of a correct angle of the needle."), open questions ("With what angle should the needle be brought in?"), feedback utterances ("That is the correct answer."), exclamations ("Do I have to explain that again?"), motivating ("Good job."), and steering utterances ("Decrease the angle."), etc. A list of dialogue functions for the tutoring process has been constructed in part based on the SWBD-DAMSL dialogue acts, where each of the functions can only be executed when several preconditions are satisfied (Jurafksy 1997; 2001) as we will explain in more detail later. Apart from the choice of dialogue act, taking into account previous dialogue acts, there is the choice of facial expressions in the tutor's face (Figure 5) and, when the tutor decides to give a demonstration of a subtask, the manipulation of objects in the virtual world.

The agent platform we use allows us to introduce new agents with different properties and communication capabilities (Hospers et al. 2003). We distinguish several types of agents: Sensor agents that are proactive and processing agents that are reactive. In the chosen domain of subcutaneous injection, the proactive "sterile agent" looks every second to see if there are objects that are not sterile when they should be. An example of a reactive agent is the "feedback agent" that acts when it receives a message from another agent; this feedback agent then will determine whether or not feedback should be given. Other agents help to observe the performing of tasks by a student and to provide explanation. These agents select their actions on the basis of what the tutor presumes to be the mental state of the student. Error-agents play an important part in this assessment.

FIGURE 5 The 3D embodied tutor agent.

From a global point of view, the three main agents used in the subcutaneous injection exercise are the TaskAnalyzerAgent, checking the correct order of the steps in the tasks and subtasks; the PhantomAnalyzerAgent, checking the speed, position, force and angle of the needle, and the location of the objects during the haptic simulation; and the SpeechAnalyzerAgent, taking care of communicating and analyzing the speech input. ErrorAgents receive information (in XML format) from these Analyzer Agents and are able to determine when something is wrong, what is wrong (angle, excessive force, exercise takes too much time, etc.), and how harmful it can be. After comparing signals coming from the haptic device with a "correct model," a message can be constructed containing relevant information for deciding whether to give instruction and the kind of instruction to be given by the tutor, using speech and facial expression. ErrorAgents that take care of the comparison with the "correct model" are alerted in critical zones around objects in the virtual environment.

The TutorAgent decides which action is appropriate action: inform the student, do nothing and let the student make an error, or decide to offer another exercise or demonstrate a subtask. Not implemented is haptic feedback that forces the student to follow a haptic path that is not too far away from the ideal path. But this is certainly an interesting option to explore in future work.

For each subtask, there are thus error agents required to monitor the student's actions and send messages to the tutor agent about these actions. Each check that has to be done is listed in Table 1. The table shows conditions for every subtask. The agents that will monitor the multimodal interaction components will send messages to the TaskAgent when the conditions as mentioned in the table are met. The agent responsible for the task analysis will deactivate the currently running error agents and will start the error agents belonging to the changed subtask. The tasks of these agents are listed with the help of (deterministic) finite automata (DFA) descriptions. In Figure 6, we display the DFA for the SterileCottonAgent. It detects errors on the disinfecting subtask of a student's exercise. The student is able to choose a sterile cotton as a representation of the avatar of the PHANToM.

In the next section, we will discuss the way this tutor agent uses affect in its interaction with the student.

The agents that currently have been implemented in the environment track the activities of the student, notice the errors that are made, interact with the student, and change the teaching environment. In particular, the earlier-mentioned ErrorAgents know about the direct performance of the student. Does the student use the right angle of the needle when trying to give an injection, what is the speed and what is the force that is used, has the needle been sterilized, does the student take too much time, does

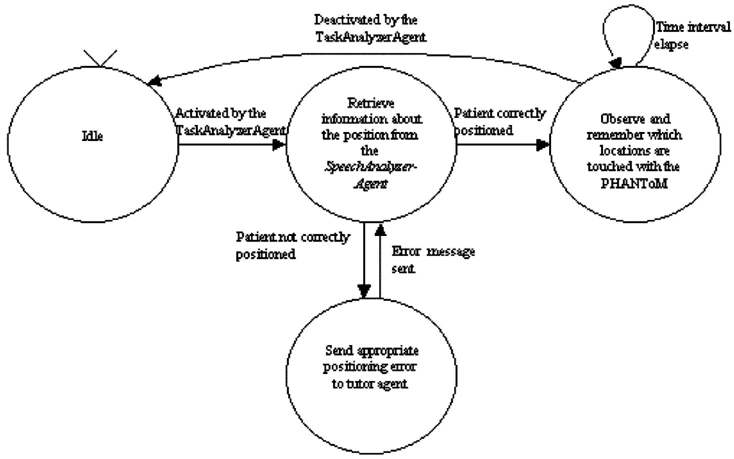

FIGURE 6 SterileCottonAgent.

he or she have many questions, or is he or she asking too often about explanations, etc. This behavior of the student and especially the perception of this behavior by the environment has to be translated to emotions that the student is assumed to have. We assume that he or she is happy about the progress that is made, unhappy because of the many errors, frustrated about not being able to master the haptic device or about the quality of speech recognition, or the lack of the teacher to understand his difficulties. Before we discuss affect in the INES system in more detail, we present the ideas on affect in tutoring that have informed our design.

AFFECT IN TUTORING

A tutor engages in interaction with a student to teach him or her certain knowledge or skills. Typical acts of the tutor include setting specific objectives for the student, motivating him or her, giving instructions, setting a specific task, asking questions, answering questions, explaining things, providing support, hinting, pumping for more information, summarizing, giving examples, providing positive or negative feedback, and evaluating the student. An important part of an intelligent tutoring consists in determining which type of action is to be carried out and the precise form this should take.

Several factors are important in this choice. First, there is the general goal of the specific lesson. The tutor may have decided on a specific strategy to get there. By choosing the Socratic method, for instance, asking questions will be selected rather than explanations. Of course, the choice of action is also determined by the kind and quality of the student's actions, i.e., how well the student does. This has effects on the mental state of the student, which should also be taken into account.

Elements of this state can be classified as relating to mood in general, feelings about motivation and achievements related to the task and attitudes towards others (the tutor, parents, and peers), and their attitudes and actions towards the student. The emotionally intelligent tutor agent that we are developing is trying to construct a model of the mental state of the student and knows about the potential effects of tutoring acts on the mental state. These insights are used to determine the appropriate action sequence and the way in which the actions are carried out. We will show how we specify conversational actions of tutors in such a way that pre-conditions and effects of the actions take the mental state of the student into account later. We will now consider the kinds of variables that one might consider in such a model.

There are four kinds of variables that we will discuss relating to the mental state of the student. These are presented in Table 2. The first row lists some of the major psychological variables that influence the way a student will carry out the task: how well is she doing, is she paying attention, what about her motivation and self-esteem, and is she attributing mistakes and successes to herself, others, or circumstances. Next, we have listed emotions that these variables will influence. The values for the emotion axis have been taken from Kort et al. (2001). They describe the cyclical dynamics of the emotions a student could or should go through during a session. We have also listed the social emotions (emotions about others or in which others play some role) that are important to consider. The latter are closely connected with some typical interpersonal relations, including dominance and affiliation, antagonism, and trust. These define how the student and the tutor relate to each other on an interpersonal level.

The elements of the mental state that have been mentioned in the table refer to the attitude towards the learning activity (achievements, attention paid to the task, motivation), the interaction with the tutoring system (collaboration, presentation), and general aspects of the student's personality that have an impact on the task (such as self-esteem). Attribution, for instance, is one of the central topics in social psychology and it plays

TABLE 2 Variables Related to Mental State

Mental state	Learning Success, Attention, Collaboration, Motivation, Self-Presentation, Self-Esteem, Attribution
Emotion axis	Anxiety – Confidence, Boredom – Fascination, Frustration – Euphoria, Dispirited – Encouraged, Terror – Enchantment
Social emotions	Embarrassment, Pride, Dislike, Joy for Other, Gratitude
Interpersonal level	Dominance, Affiliation, Trust

a crucial part in the tutoring situation. It concerns the way we explain our own behavior and that of others. In this case, attributions typically concern the success or failure of the actions of the student. Typically, tutors and students will disagree about certain attributions. Perceivers tend to overestimate personal or dispositional factors and underestimate situational factors. Actors, on the other hand, might attribute their actions more to situational factors.[3] Such different biases (also related to personality traits, which are not mentioned in the table) have been investigated in quite some depth.

Attributions will affect other parameters of the mental state, including motivation and emotion. Relevant in this respect is, for instance, the work by Weiner (1995) who made a distinction between two kinds of achievement-related affects: *outcome-dependent* (did we achieve our goals or not) and *attribution-linked* affects (was it because of what we did or not). This means that events influence the emotional state in two stages, relating different elements of Table 2: failure leads to outcome-dependent affect (misery, embarrassment), and causal attribution of the failure (self/other/circumstances) leads to future expectancies (hope/no hope) and associated affect (depressed/optimism). These kinds of effects will have to be considered by the emotionally intelligent tutor.

If we want to build tutor agents that choose their actions and the way they perform them, taking into consideration the emotional state of the student, then it is important to ask ourselves what emotional states a tutor should want to achieve. In Kort et al. (2001), it is explained how the learning cycle that learners experience is matched with a model of emotions. The emotions that students go through are described in a cyclic movement. In emotional terms, the motivation to solve a problem may arise out of curiosity (a positive incentive) or puzzlement and confusion. This means that negative emotions may have positive effects for the learning process. In the choice of action (the kind, manner, sequence), tutors will have to take such values into account and through their actions try to change them in the direction desired.

The emotional state of the student contributes a lot to whether a student is motivated or challenged, which are key conditions for certain actions. Curiosity and puzzlement may lead to investigate problems. But also frustration may lead to action, even though it is a more negative affect. The tutor can choose to consider taking certain actions to bring about a change in the emotional state. Lepper (1993) identified four main goals in motivating learners: challenge them, give them confidence, raise their curiosity, and make them feel in control. These goals can be achieved by means of various tactics. The student can be challenged by selecting appropriately difficult tasks, or by having the difficulty emphasized or by having some kind of competition set up. Confidence can be boosted by maximizing success directly (praising) or indirectly ("it was quite a difficult task, you managed to do"). Curiosity is typically raised in Socratic methods when the student is asked to ponder many questions. The tutor can decide to leave the initiative to the student or offer options that suggest the student can make choices and thereby influence the student's feeling of being in control. Lepper's goals have been taken into account in our system in the choice of learning component or teaching strategy (Socratic methods, active student learning, using deep explanatory reasoning, etc.).

It is clear that, in general, the affective state of the student is important during the learning process. The values for the emotions in the emotion axis cell refer specifically to how the student feels about the problem she is facing in the learning process and how she can cope with that. But given the fact that there are others the student is interacting with, like the tutor, her peers, and parents, social emotions should be taken into account as well. The entire teaching process consists in negotiating about obligations to do certain tasks and evaluations of the result. Almost every action carried out by the tutor has to take such variables into account. Note that the social emotions, again, will depend to a large extent on attribution of the cause of an action on oneself or others.

Intricately connected with the social emotions are the interpersonal attitudes. Two important factors that are often discussed are dominance and affiliation. Also, in the case of a student-tutor relation, these factors play an important role. The type of relationship that a tutor wants to establish with the student will be defined to a great extent by these parameters. The degree of dominance and affiliation can be varied and this will clearly have an impact on the kinds of actions that are chosen or the teaching style in general. To see the relevance of these variables to the tutoring situation, one can look at the following Table 3 taken from Argyle (1972).

Here, Argyle lists some typical actions or techniques that are often used in these different strategies. These acts were not specifically introduced for a tutoring context. However, it is immediately obvious that many actions of tutors are specific instances of the actions in the table: analyze, criticize, disapprove, advise, lead, agree, etc. The choice of action will correspond to a specific teaching style and define in part the interpersonal relation between tutor and student. There are several other interpersonal actions that give rise to affect appraisals. Consider, for instance, the difference between the psychological effect of formulating a task as an order or as a suggestion. The tutor has to steer and motivate the student, know when the student welcomes a hint, etc. In general, one of the goals that people want to come out of social interaction is to enhance the self of each actor. The ideal outcome is that the student is proud of his achievements and feels highly esteemed by the tutor. The emotional state related to this form of social

TABLE 3 Interpersonal Attitudes and Actions

	Dominance		
Affiliation	Analyzes Criticizes Disapproves Judges Resists	Advises Coordinates Directs Leads Initiates	**Affiliation**
low	Evades Concedes Relinquishes Retreats Withdraws	Acquiesces Agrees Assists Cooperates Obliges	**high**
	Dependency		

interaction typically involves elements and variables such as: social rewards, dependence, status, power, and face.[4]

We will show next how our work involves specifying and classifying such tutor actions, as specific conversational moves. The tutoring situation is primarily a dialogue, and therefore insights from social psychology related to these variables will have to be incorporated in this specification of the dialogue acts involved. Affective parameters in tutoring dialogues not only concern the choice of conversational act and the overall tutoring strategy, but also the kind of language that is used. In tutoring dialogues, we see a number of specific dialogue acts characteristic of tutoring. They make tutoring conversations different from other types of conversations, say information dialogues and task dialogues. Some of the dialogue acts are special for affect control, either for controlling the atmosphere of the conversation or for controlling the motivational state of the student. This is an immediate consequence of the specific roles, goals, and tasks the participants have in a tutorial conversation. The student's goal is to learn something, to solve a theoretical problem, or to master a practical task. The task of the tutor is to encourage the student's active learning behavior by guiding him, giving hints, and asking questions to stimulate the student to reflect on what he is doing, carefully avoiding to give away the solution to the problem. Types of responses that are completely unsatisfactory in other types of conversations are common in tutorial dialogue. Tutorial dialogue shows extensive use of acknowledgment acts and accept acts by the tutor—showing his understanding and encouragement—and a lack of explicit signals from the student's side that he will cooperate. In case the learner has to master a complex practical task, the tutor has to follow the student's actions, to see whether he is doing the subtasks in the correct order and whether he is doing the subtasks well. The tutor has to evaluate the student's actions and give appropriate feedback, positive or negative, depending on the student, and the complexity of the task. When the student makes an error, he can choose to interrupt the student or to leave him and wait until the student finds out himself. The tutor has to see whether the student makes enough progress in learning the task, and, if not, to figure out what are the causes of the mistakes he makes. In all of these tasks, the tutor should have a good feeling for the affective and motivational state of the student.

AFFECT IN INES

Our aim is to make an intelligent tutoring system that builds on the psychology of affect and emotion to make the learning process more efficient. When we view the tutoring situation as a dialogue, the important question becomes: What should the tutor say and do in a particular situation to optimize the learning curve?

It seems less important that tutors show their own real emotions. But we think that tutors should show empathy with the students and understand how their emotional state is affected by what the tutors do. In Gratch et al. (2002), Gratch and his co-authors make a distinction between communication-driven approaches and simulation-based approaches to modeling emotion and personality for building believable virtual humans. "In communication-driven approaches, a virtual human chooses its emotional expression on the basis of its desired impact on the user .... Tutoring applications usually also follow a communication-driven approach, intentionally expressing emotions with the goal of motivating the students and thus increasing the learning effect .... The second category of approaches aims at simulation of 'true' emotion (as opposed to deliberately conveyed information)." Cosmo (Lester et al. 2000) is cited as an example of the first approach, whereas in Marsella and Gratch (2001), the second approach is taken. On the one hand, our approach focuses on the communication between the tutor agent and the student, but only in this sense can we consider it a communication-driven approach. Our specific goal in creating a socially or emotionally intelligent agent, however, is not to build an "emotional agent" that can *express* its emotions, but rather an agent that chooses its actions (kind and manner) appropriately, in accordance with the mental state of the student. Of course, this could have as a side effect that the tutor is considered to show great empathy with the student. And, in this way, the tutor can also be said to show emotions.

Affect in INES is considered from several angles and on different levels. First, there is the emotional state of the student that will be modeled by an event appraisal system. Second, the emotional state of the tutor is modeled as well, including values for emotions and parameters such as satisfaction, disappointment, and surprise. Third, the dialogue acts come in different forms, with variation in affective values. And, fourth, various affective parameters are used in determining which tutoring strategy to use and which instructional act to perform (sympathizing or nonsympathizing feedback, motivation, explanation, steering, etc.).

The INES system contains emotion appraisers. These set the intensity of a few emotion types based on the events that occur. There are two sets of values. One set models the tutor emotions and another set models the student emotions. These values are taken into account in determining the nature and the form of the dialogue act that will be performed.

In order to make a good guess of how the student's affective system reacts in particular situations, the tutor also has to have some idea of the character of the student. This may refer to such properties as being introverted, self-critical, extroverted, anxious to do something wrong, or self-confident. A believable tutor agent needs to monitor the emotional and motivational state of the student. Recognition of the affective and emotional state of the student is a difficult task for a tutoring system. It assumes automatic affect recognition in text and prosodic properties of speech, in facial expressions, and in other aspects of nonverbal behavior. Moreover, we need ways to automatically detect features to induce the state of concentration, motivation, and engagement of the student. Instead of trying to determine the emotions of the real student by means of affective computing techniques, the ITS could also build a model of the student emotions by using appraisal rules. This is the route that we have pursued for the current version.

The main elements that go into calculating the emotional state of the tutor are: the level of activity of the student, the difficulty of the task, the history of the student (how well he/she did in previous tasks), the number of errors made, and the harmfulness of the error. For the tutor model, we have limited our prototype implementation to four values, representing the emotions *joy, distress, happy for,* and *sorry for.* With respect to the student model, we have taken into account, so far, four emotional categories that affect the selection of the kind and form of instruction. These are *anxiety–confidence* and *dispirited–enthusiasm*.

The central part[5] of INES that will be responsible for addressing the issues addressed in the previous section is the instruction module that decides on the sequence of instructions that will be given. This is shown in Figure 7.

In working with INES, students start by choosing an exercise to perform (see Table 4). The system uses descriptions of exercises that contain information about the different subtasks involved (partially ordered), the kinds of errors, pointers to error-agents that can detect the errors, error messages, and steering acts, explanations, hints, questions, and answers. Many of these are used to specify the precise output of INES. The output consists of text messages, spoken output, expressions on a 3D face, and haptic feedback through the phantom. The main conversational and tutoring functions that have to be performed are: make conversation (welcome, introduction, closing), give instruction, set a task, answer questions, ask a question, provide support, give feedback, explain, evaluate, motivate, expect, and steering the student along. In a typical sequence, the tutor presents a problem to the student, the learner starts to solve the problem, the tutor assesses the answer and provides feedback. Such sequences often loop and have alternative branches: in cases of

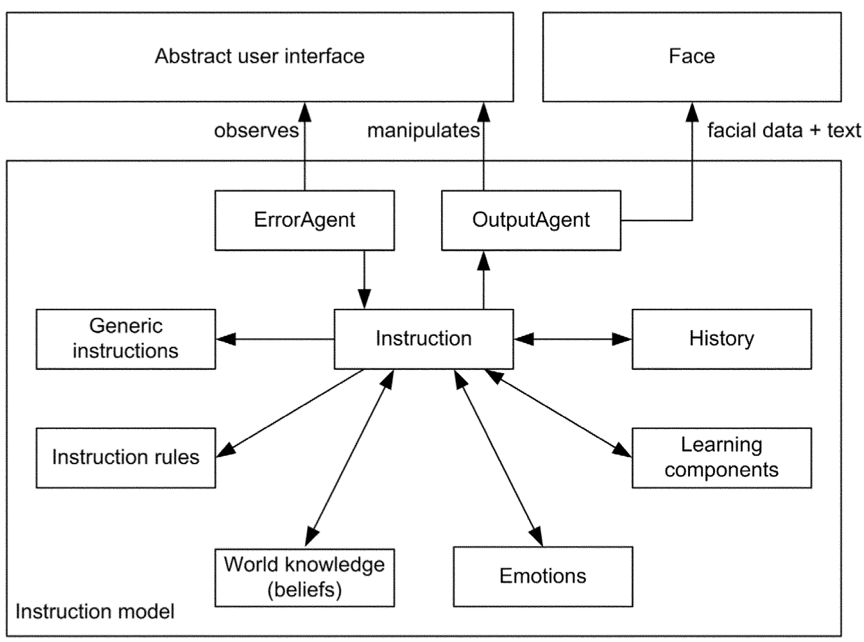

FIGURE 7 Overview of the different agents and components in the tutor component.

TABLE 4 Description Exercise File Elements

Element name	Description
exercisefile	The exercisefile contains zero or more exercises.
Exercise	An exercise.
Settings	A pathname to the class containing all information about the environment. This information includes physical object information.
ExerciseID	A unique exercise ID.
Phase	The phase the student should be in to be able to do this exercise.
Description	A short description of the exercise.
Subtask	A subtask.
SubtaskID	A unique identifier (in this exercise) for the subtask.
ErrorAgent	An ErrorAgent to use in this subtask. An ErrorAgent should have a name and a link that contains the path where the agent can be found.
SubtaskDescription	A short description of the subtask.
Difficulty	The difficulty of the subtask.
MaximumTime	The maximum time the student can work at this subtask.
Error	A specific type of error the student can make. This element needs to have a name (the type of error) and an agent's name (corresponding with an ErrorAgent defined) from which the error types will be received (e.g., InjectionAgent).
Examples	Examples that can be demonstrated to the student.
Questions	Questions that can be asked to the student. Questions can be specific for an error type or globally defined for the total subtask.
Harmful	Indicates in what way an action is harmful. It can have an error kind (as a precondition) attached to it (e.g., "*tooDeep*") to be able to specify the harmfulness for a specific error kind.
Evaluation	Evaluation texts that can be given. It should have a precondition to be able to evaluate each specific error kind.
explanation_hints	The explanations and hints that can be given.
Steering	The steering acts that can be used.
expl_hint_type	A type of explanations and hints. It has a type attached to it that says what kind of explanation or hint it is (e.g., it gives away the correct answer, or it emphasizes the error).
Pre	A precondition counts for several explanations and hints. It should have an error kind attached to it.
Explanation	An explanation that can be given.
Hint	A hint that can be given.
Steer	A steering act that can be given. It should have an error kind attached to it.
question_type	Contains types of questions that can be asked. It should have a type as a precondition indicating what kind of questions are contained (multiple choice, prompting, open, yes/no).
question_answer	Contains a question and its answer.
Question	A question that can be asked.
Answer	An answer to a specific question.
Example	An example that can be given.

misunderstanding, right or wrong answers, etc. Given a specific choice for a teaching strategy, certain actions will be preferred. The major question to be answered in the current project is which action and strategy should be chosen and how are these to be performed, taking the factors discussed into account.

The module termed "learning components," decides on what kind of learning component (see Graesser et al. [in press]) or teaching strategy is taken. For instance, if the Socratic method is active, the tutor will tend to ask more questions. The choice of learning component can be changed dynamically during a tutoring session. The choice is mainly based on the activity level of the student and the particular stage in an exercise. In the current prototype implementation, we have considered active student learning, deep explanatory reasoning, and affect and motivation as learning components, and the Socratic method as a pedagogical strategy. The error-agents, which we discussed earlier, provide the instruction module with information about how the student is performing the task. The kind of error students make, the harmfulness of the error, and the number of times a student has made it are the main input to determine the kind of response.

The selection of the strategy depends mainly on the activity level of the student. Three levels are introduced on which the tutor can operate. The idea is that there is congruence between the student's self-steering and the tutor's external steering. Thus, on level 1, when the student isn't active at all, fixed external steering is used. The tutor dynamically switches between the different levels, taking into account the performance of the student. The idea of using different levels of activity is based on an algorithm in Fiedler (2003), which determines what kind of hint (with what kind of activity) should be given, taking into account the student's current and previous answers, the number of wrong answers, the previous hints produced, and the number of hints produced. In our prototype, we start at the highest level. That means that we want the student to be active and the tutor interventions are kept minimal.

- Start in the initial level. This will be as active as possible, so this is level 3 (the student file can be adapted by adding an initial level).
- When the student makes an error, the tutor responds to this error and doesn't do anything yet.
- When the student makes the same error again (same error type and error kind), adjust the level to a lower one and use it in the response.
- When the student does OK on all error types for a while, the level is increased.

The choice of strategy influences the kinds of actions that are performed and the sequence in which they are performed. For instance, when the Socratic method is active, the tutor will choose to ask a lot of questions. We are interested in having our tutoring system teach the students certain practical skills with an insight into the reasons for doing tasks in a specific way. An important part of tutoring sessions is, therefore, to let the students practice some task. In this case, the tutor simply observes what the students does and provides feedback. Ideally, the student should be active and self-motivated to practice the tasks.

Because the typical exercises involve practicing a task, the error-agents form an important component of the system. They observe the student's actions, diagnose them, and report to the instruction-agent: the multi-modal dialogue manager or brain behind the ITS that determines the sequence of instructions. For instance, the sequence ⟨feedback, motivation, support⟩. This is called the teaching task *support_tt* and is typically instantiated in the case of an haptic error (see Table 5).

TABLE 5 Teaching Tasks

Type	Instruction rules
Haptic error	⟨support_tt⟩ = feedback (motivation | exclamation) support
	⟨back_question_tt⟩ = feedback question_contents
	⟨explain_tt⟩=feedback exclamation explanation set_task
	⟨example_tt⟩=feedback exclamation explanation example
	⟨steer_tt⟩=steering
Haptic correct	⟨agree⟩= feedback_student
Subtask intro	⟨introduction_tt⟩=conversation expectation set_task
Exercise finished	⟨close_tt⟩=feedback evaluation error_enumeration conversation
Exercise intro	⟨welcome_tt⟩=conversation

For each of the instructions, called dialogue tasks, a dialogue function will have to be chosen. These are more specific instances of the dialogue tasks. For support, there are functions such as "hint correct answer," "hint error emphasize," and "hint error diagnosis." For each dialogue function there are several specific dialogue acts. These can be verbal or nonverbal messages (head nods, for instance), or feedback given through the haptic device.

A major part of our work takes the form of a formal specification of the different dialogue moves of a tutor. Typical dialogue actions are redefined by subtyping and parameterizing them further. Pre- and postcondions in these specifications will relate to the variables discussed earlier For instance, whether or not and how a tutor sets a task, and the kind of task (difficult, easy) for a student can depend on the strategy chosen, on the motivational state of the student, or her achievements so far. The acts can be further distinguished along the manner in which they are executed, where manner depends, amongst others on the interpersonal relationship. The specification of the different tutoring acts are similar to the definitions of speech acts as provided by Searle (1979), indicating the different preparatory conditions and the intended effects. It is interesting to recall some of the dimensions of variation (12 in total) that Searle uses to distinguish illocutionary acts from one another. Searle claims that there are differences in the point (or purpose) of the (type of) act, in the expressed psychological states, in the force or strength with which the illocutionary point is presented, in the status or position of the speaker and hearer as they bear on the illocutionary force, in the way the utterance relates to the interests of the speaker and the hearer, and in the style of performance of the illocutionary act. It is immediately obvious that these dimensions that Searle refers to directly relate to the variables we are interested in.

To illustrate how this works we provide a simple example. The following is a generic definition for a general set_task act that does not refer to any of the variables discussed above.

set_task_act(T,Task,S)

Constraints:	tutor(T) & task(Task) & student(S) & knowledge(K) & example(E)
Precondition:	know(T,Task) & want(T,set_task_act(T,Task,S) & (know(S,related(Task,E)) or know(S,related(Task,K)))
Effect	do(S,Task)

In our socially intelligent variation on this definition, this is further refined. The conditions referring to the social variables are similar to Searle's *sincerity conditions*. Setting a task for a student is a kind of "directive." It can be put forward as a suggestion, a puzzle, or an assignment. Depending on the past achievements, the difficulty of the task can be adjusted. The task can be provided to set a challenge for the student and so it can change the emotional state of the student. If the student is set a task, then the intended effect is that the student will perform the task. Whether or not the student will perform the task will, of course, depend on the motivational state. Table 6 shows how some emotional parameters determine in part the selection of a specific feedback act. In Figure 8 the structure of the instructions is shown.

TABLE 6 Partial Dialogue Function Determination in Words (Feedback)

Feedback

- Feedb_part_reject:
  - If satisfaction and it is an error (intensity bigger than 3).
  - If disappointment and a small error or 3.
- Feedb_reject:
  - If disappointment and it is a big error.
- Feedb_part_accept:
  - If disappointment and not an error and not an intensity of 3.
- Feedb_accept: If satisfaction and it is not an error.

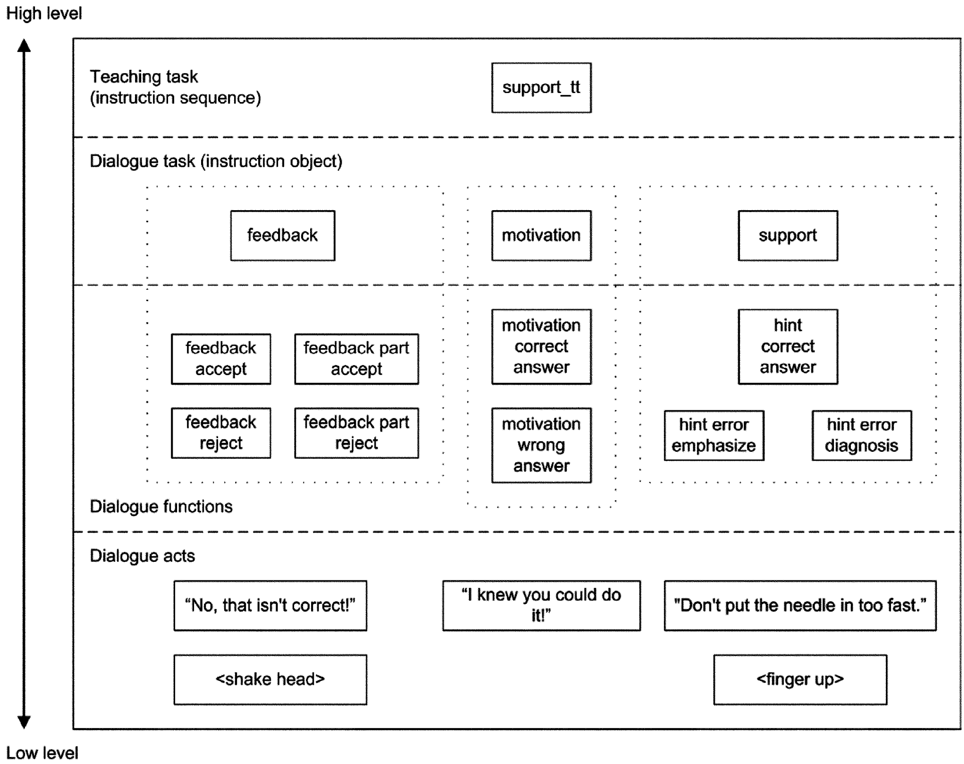

FIGURE 8 Structure of instructions.

Although we have not yet performed an evaluation with users, we have carried out several types of tests of the system. We have tested the effect of various parameters on the choice of dialog acts by letting the system respond to the same input with different settings for these parameters. The behavior of the student was simulated according to specific scenarios. The tests were conducted to see how the tutor's conversational responses and his emotion expression varies with the settings of a number of parameters in a scenario in which the student repeatedly makes the same error, namely, he brings in the needle at an angle that is too wide. The tests were performed in varying situations in which the following features were controlled.

1. The tutor shows his own emotion or not. Four emotion categories were implemented and tested for their effect on the conversational behavior: *joy, distress, happy for,* and *sorry for.*
2. The tutor takes the student's emotions into account,[6] whether the student is either *confident* or *insecure.*

3. The tutor follows a Socratic learning strategy and stimulates active student learning. The tutor then favors asking questions above giving lengthy explanations.
4. The harmfulness of the error the student makes differs.

We present two dialogues between the emotional tutor with a confident versus an insecure student, respectively, to illustrate the differences in the responses that arise with the same kind of input.

A dialogue between a confident student and an emotional tutor:

[S] Uses an angle that is too wide.
[T] Shows a negative facial expression (sadness).
[S] Uses an angle that is too wide.
[T] Not quite. It's okay. Take care of the angle with which you put the needle in.
[S] Uses an angle that is too wide.
[T] Stop. That's wrong! You shouldn't inject with an angle that is too wide. Try it again.
[S] Uses an angle that is too wide.
[T] Try again. I will steer you along this time.
[S] Uses an angle that is too wide.
[T] Decrease the angle.

A dialogue between an insecure student and an emotional tutor:

[S] Uses an angle that is too wide.
[T] Shows a negative facial expression (sadness).
[S] Uses an angle that is too wide.
[T] Not quite. You use a wrong angle. Okay, start again.
[S] Uses an angle that is too wide.
[T] Not quite. It doesn't matter. Take care the angle isn't too wide.
[S] Uses an angle that is too wide.
[T] Wait a minute, you're in the wrong direction. It's okay. Try to keep an angle of thirty to forty-five degrees.
[S] Uses an angle that is too wide.
[T] Not quite. I will give a demonstration of the task.

These two dialogue fragments clearly show that setting the parameters differently causes significant differences in the teaching tasks and dialogue functions. In the second case, the tutor responds in a more sympathetic way, using more motivational verbal instructions, no exclamations, and no total rejection. We have let different people judge the various scenarios (showing them not just how the dialogue unfolded, but also the facial expressions) and, in general, the output differs along with what people would expect to change in the various situations. Of course, the dialogue is contrived, because the student keeps making the same error. Also, the prototype uses several responses that may seem a bit awkward or exaggerated. This can be fixed by adjusting the exercise files and is, therefore, a matter of scripting. We will need to fine tune one or more exercises before we can carry out further evaluations of the system.

CONCLUSIONS

We have presented INES as an intelligent tutoring system that takes into account affective parameters related to the tutoring situation. These parameters determine the course of the dialogue. This shows our perspective on the design of emotional embodied conversational agents. We want to emphasize in particular their social and interpersonal functioning. In this paper, we have provided some background motivation for this approach and looked at ongoing work in specifying emotionally intelligent tutors. We have reviewed the major parameters that should be considered in the development of such agents. Conversation constitutes the major part of the tutoring activity and communication is a social act. This means that in developing socially intelligent tutors it is very important to define social intelligence in terms of the conversational acts that take place. We have indicated how this can be done by letting the choice of teaching strategy or the definition of specific conversational acts depend on emotional and interpersonal variables.

We stated in the introduction that building socially intelligent agents requires thinking of them as agents that can construct theories about the beliefs and emotions of the person or other agents in which they interact. An important issue thus involves determining the mental state of the student. Obviously, not all of the parameters are easy to monitor. There are different ways to get information. Estimates of what the student is thinking or feeling with respect to the situation can partly be made on the basis of general models and patterns, i.e., on typical expectations of how people behave. The way the student actually responds can be taken into account to adjust this information to the particular situation. Some of the things that can be measured are the time the student takes to answer (hesitation/distraction), the correctness of the answers, etc. Certain information can also be obtained by explicitly asking the student questions related to their mental state. Information from previous interactions can also be taken into account. These are the kinds of cues that we have relied on in our systems. One could also consider taking physiological measurements into account or try to find out how the student feels by analyzing her facial expressions, for instance. We are currently considering some of these forms of affective computing. Building conversational agents that can adequately read our minds will remain an interesting challenge for some time, but we think that to model social interaction this will be necessary at least to some degree.

NOTES

1. There are many research projects on haptics and computer graphics (virtual reality) for medical training. They aim to allow students to train on virtual patients rather than having to practice on a vulnerable human patient. Among them are projects on injection simulators, where trainees feel realistic forces when attempting to position the needle and inject a fluid and where different feedback can be obtained for different vein sizes and patient profiles, (e.g. Dang et al. [2001]). Unlike what has been done in the previously mentioned systems, in our research we aim at the interaction between student and system, between student and tutor, and between student and patient, allowing integrated multimodal interaction, and where haptics is only one of the modalities that is used.
2. There is a limited build-up of the interaction history. A more complete interaction history will be obtained by embedding the characteristics of a generic multimodal interaction architecture (see our work reported in Hofs et al. [2003]) in the INES environment. This earlier work contained models of multimodal interaction—in that particular case tuned to speech and pointing—that can be used for haptic input combined with speech, mouse, and keyboard input, and that allows the embedding of multimodal input in a discourse model that also keeps track of the history of the interaction.
3. In our case, if a task does not succeed, this may be because the student makes an error, but also because the system does not function without flaws. The speech recognition, for instance, regularly makes mistakes beyond the student's control.
4. In Johnson et al. (2004), politness theory and face management are considered in generating appropriate dialogue moves for a tutor agent.
5. More information about INES can be found in Hospers et al. (2003) and Vissers (2003).
6. These states correspond with two quadrant spaces: anxiety vs. confidence and dispirited vs. enthusiasm of Kort et al. (2001).

REFERENCES

Argyle, J. M. 1972. *The Psychology of Interpersonal Relations*, 2nd ed. London: Penguin.

Baron-Cohen, S. 1995. *Mindblindness. An Essay on Autism and Theory of Mind.* Cambridge, MA: The MIT Press.

Bui, T. D., D. Heylen, M. Poel, and A. Nijholt. 2002. ParleE: An adaptive plan-based event appraisal model of emotions. In *Advances in Artificial Intelligence, Proceedings KI 2002, 25th German Conference on Artificial Intelligence*, eds. M. Jarke, J. Koehler, G. Lakemeyer, pages 129–143, Lecture Notes in Computer Science 2479. Berlin, Heidelberg, New York: Springer.

Burghouts, G. J., R. op den Akker, D. Heylen, M. Poel, and A. Nijholt. 2003. An action selection architecture for an emotional agent. In *Recent Advances in Artificial Intelligence, Proceedings of FLAIRS 16*, eds. Russell and Haller, pages 293–297. Menlo Park: AAAI Press.

Dang, T., M. A. Srnisvasan, and T. M. Annaswamy. 2002. Development and evaluation of an epidural injection simulator with force feedback for medical training. In *Proceedings of Medicine Meets Virtual Reality Conference* 10, pages 97–102, Newport Beach, CA, USA.

Fiedler, A. and D. Tsovaltzi. 2003. Automating hinting in mathematical tutorial dialogue. In *Proceedings of the EACL 2003 Workshop on Dialogue Systems: Interaction, Adaptation and Styles of Management*, pp. 45–52, April, Budapest, Hungary.

Goleman, D. 1997. *Emotional Intelligence.* New York: Bantam Books.

Graesser, A. C., N. K. Person, and D. Harter. Teaching tactics and dialog in AutoTutor. *International Journal of Artificial Intelligence in Education.* in press.

Gratch, J., J. Rickel, E. André, N. Badler, J. Cassell, and E. Petajan. 2002. Creating interactive virtual humans: Some assembly required. *IEEE Intelligent Systems.* 17(4):54–63.

Heylen, D. 2003. Facial expressions for conversational agents. In: *Subtle Expressivity for Characters and Robots: CHI 2003 Workshop*, eds. N. Suzuki and C. Bartneck (www.bartneck.de/workshop/chi2003).

Hofs, D., R. op den Akker, and A. Nijholt. 2003. A generic architecture and dialogue model for multimodal interaction. In *Proceedings of the 1st Nordic Symposium on Multimodal Communication*, eds. P. Paggio, K. Jokinen, A. Jönsson, pages 79–92, Copenhagen, Denmark.

Hospers, M., E. Kroezen, R. op den Akker, A. Nijholt, and D. Heylen. 2003. Developing a generic agent-based intelligent tutoring system. In *Proceedings of the International Conference on Advanced Learning Technologies*, page 443, Athens, Greece.

Hospers, M., E. Kroezen, A. Nijholt, R. op den Akker, and D. Heylen. 2003. An agent-based intelligent tutoring system for nurse education. In *Applications of Intelligent Agents in Health Care*, eds. J. Nealon and A. Moreno, pages 143–159. Whitestein Series in Software Agent Technologies. Basel: Birkhauser Publ. Ltd.

Johnson, L. W., P. Rizzo, W. Bosma, S. Kole, M. Ghijsen, and H. van Welbergen. 2004. Generating socially appropriate tutorial dialog. In *Affective Dialogue Systems*, eds. E. André, L. Dybkjaer, and W. Minker, pages 254–264. Berlin, Heidelberg, New York: Springer Verlag.

Jurafsky, D. 1997. *Switchboard SWBD-DAMSL Shallow-Discourse-Function Annotation.* Technical Report 97-02, University of Colorado.

Jurafksy, D. 2001. Pragmatics and computational linguistics. In *Handbook of Pragmatics*, eds. L.R. Horn and G. Ward, pp. 578–604. Oxford: Blackwell.

Kort, B., R. Reily, and R. Picard. 2001. External representation of learning process and domain knowledge: Affective state as a determinate of its structure and function. In: Proceedings AI-ED 2001 Workshop on External Representations in AIED: Multiple Forms and Multiple Rows, May, San Antonio, Texas.

Lepper, M. 1993. Motivational techniques of expert human tutors: Lessons for the design of computer-based tutors. In *Computers as Cognitive Tools*, eds. S. Lajoie, and S. Derry, 75–105. Hillsdale: Lawrence Erlbaum Associates.

Lester, J. C., S. G. Towns, C. Callaway, J. L. Voerman, and P. J. Fitzgerald. 2000. Deictic and emotive communication in animated pedagogical agents. In: *Embodied Conversational Agents*, eds. J. Cassell, J. Sullivan, S. Prevost, and E. Churchill. 123–154. Cambridge, MA: The MIT Press.

Marsella, S. and J. Gratch. 2001. Modeling the interplay of emotions and plans in multi-agent simulations. In *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*, pages 622–627. Edinburgh, Scotland, August.

Picard, R. W. 2001. An affective model of interplay between emotions and learning: Reengeniering educational psychology - building a learning companion. In *Proceedings of the International Conference on Advanced Learning Technologies (ICALT)*, pages 43–48, Madison, WI, USA.

Searle, J. R. 1979. The classification of illocutionary acts. *Language in Society.* 5: 1–24.

Vissers, M. 2003. *Social and Strategic Intelligent Tutors.* Master's Thesis, University of Twente.

Weiner, B. 1995. *Judgments of Responsibility: A Foundation for a Theory of Social Conduct.* New York: Guilford Press.