Expressive Behaviors for Virtual Worlds

Stacy Marsella[1], Jonathan Gratch[2], and Jeff Rickel[1]

[1] USC Information Sciences Institute
4676 Admiralty Way, Marina del Rey, CA 90292
marsella@isi.edu, rickel@isi.edu
[2] USC Institute for Creative Technologies
13274 Fiji Way, Marina del Rey, CA 90292
gratch@ict.usc.edu

Summary. A person's behavior provides significant information about their emotional state, attitudes, and attention. Our goal is to create virtual humans that convey such information to people while interacting with them in virtual worlds. The virtual humans must respond dynamically to the events surrounding them, which are fundamentally influenced by users' actions, while providing an illusion of human-like behavior. A user must be able to interpret the dynamic cognitive and emotional state of the virtual humans using the same nonverbal cues that people use to understand one another. Towards these goals, we are integrating and extending components from three prior systems: a virtual human architecture with a wide range of cognitive and motor capabilities, a model of task-oriented emotional appraisal and socially situated planning, and a model of how emotions and coping impact physical behavior. We describe the key research issues and approach in each of these prior systems, as well as our integration and its initial implementation in a leadership training system.

1 Introduction

A person's emotional state influences them in many ways. It impacts their decision making, actions, memory, attention, voluntary muscles, etc., all of which may subsequently impact their emotional state (e.g., see [2]). This pervasive impact is reflected in the fact that a person will exhibit a wide variety of nonverbal behaviors consistent with their emotional state, behaviors that can serve a variety of functions both for the person exhibiting them as well as for people observing them. For example, shaking a fist at someone plays an intended role in communicating information. On the other hand, behaviors such as rubbing one's thigh, averting gaze, or a facial expression of fear may have no explicitly intended role in communication. Nevertheless, these actions may suggest considerable information about a person's emotional arousal, their attitudes, and their focus of attention.

Our goal is to create virtual humans that convey these types of information to humans while interacting with them in virtual worlds. We are interested in virtual worlds that offer human users an engaging scenario through which they will gain valuable experience. For example, a young Army lieutenant could be trained for a peacekeeping mission by putting him in virtual Bosnia and presenting him with the sorts of situations and dilemmas he is likely to face. In such scenarios, virtual humans can play a variety of roles, such as an experienced sergeant serving as a mentor, soldiers serving as his teammates, and the local populace. Unless the lieutenant is truly drawn into the scenario, his actions are unlikely to reflect the decisions he will make under stress in real life. The effectiveness of the training depends on our success in creating engaging, believable characters that convey a rich inner dynamics that unfolds in response to the scenario.

Thus, our design of the virtual humans must satisfy three requirements. First, they must be believable; that is, they must provide a sufficient illusion of human-like behavior that the human user will be drawn into the scenario. Second, they must be responsive; that is, they must respond to the events surrounding them, which will be fundamentally influenced by the user's actions. Finally, they must be interpretable; the user must be able to interpret their response to situations, including their dynamic cognitive and emotional state, using the same nonverbal cues that people use to understand one another. Thus, our virtual humans cannot simply create an illusion of life through cleverly designed randomness in their behavior; their inner behavior must respond appropriately to a dynamically unfolding scenario, and their outward behavior must convey that inner behavior accurately and clearly.

This paper describes our progress towards a model of the outward manifestations of an agent's cognitive and emotional state. We review three prior systems that have heavily influenced our thinking on expressive behaviors, discussing the unique aspects of each and illustrating how they have influenced the design of an integrated system. The first, Steve [40, 42, 41], provides an architecture for virtual humans that can collaborate with human users and other virtual humans in 3D virtual worlds. Although Steve did not include any emotions, its broad capabilities provide a foundation for the virtual humans towards which we are working. The second, Jack and Steve, provides a model of how emotions arise from the relationship between environmental events and an agent's plans and goals [16], as well as a model of socially situated planning that builds on that emotional appraisal model. The third, Carmen's Bright IDEAS [31], contributes a complementary model of emotional appraisal as well as a model of the impact of emotional state and coping on physical behavior. After describing the key concepts in each of these prior systems, we describe a new project in which we have integrated these concepts into virtual humans for experiential learning in engaging virtual worlds.

2 Steve

Our earliest work on virtual humans resulted in Steve (Figure 1), an animated agent that collaborates with human users and other virtual humans on tasks in 3D virtual worlds [40, 42, 43]. Such task-oriented collaboration requires an agent to balance a variety of demands. Tasks require an agent to perceive the state of the virtual world, assess the state of goals, construct plans to achieve those goals, navigate through the virtual world and execute its plans. Collaboration requires these task-related behaviors to be interleaved with face-to-face social interactions with others (human users and virtual humans) embedded in the same virtual world. The agent's environment is unpredictable in many ways: others may speak to the agent or take actions in the world at any time, and the virtual world itself may change unexpectedly (e.g., through simulated equipment failures). Thus, the agent must be able to adapt its task-related and social behaviors at any time. Steve's main contribution is its ability to interleave task-related behaviors and face-to-face dialogue in such dynamic virtual worlds.

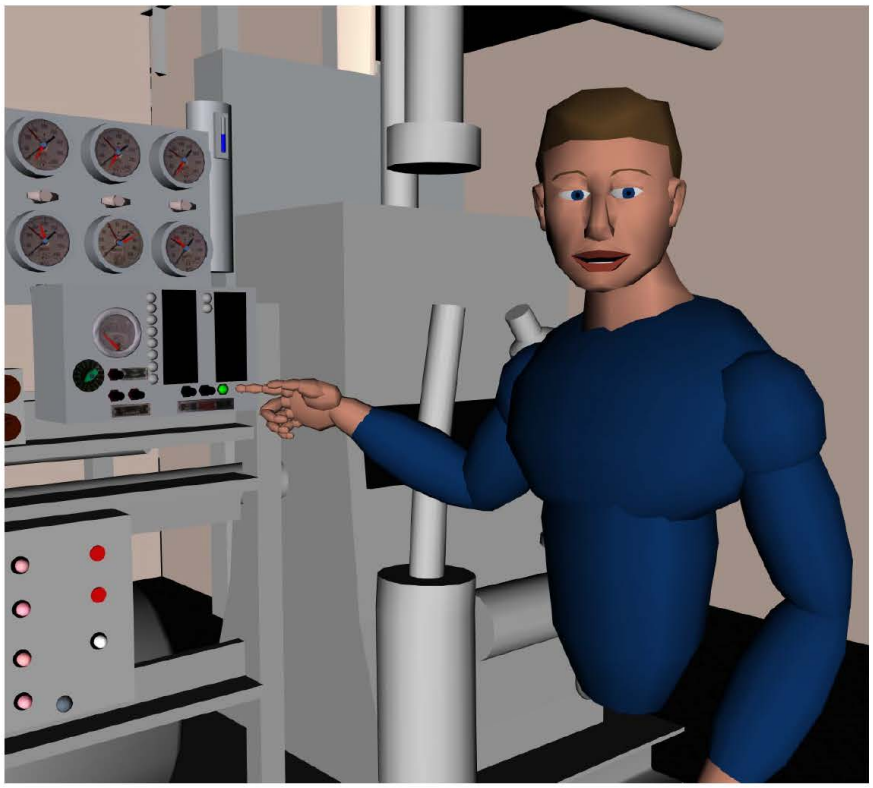

Fig. 1. Steve describing a power light

Despite an impressive variety of related work on embodied conversational agents [6] and animated pedagogical agents [21], Steve is unique in this ability. Several systems have carefully modeled the interplay between speech and nonverbal behavior in face-to-face dialogues [5, 7, 4, 36], but these systems focused exclusively on dyadic conversation, and they did not allow users and agents to collaborate on tasks in a 3D virtual world. The Gandalf system [7], which provided a sophisticated model of real-time face-to-face interaction, allowed an agent and human to cohabit a real physical space, and to use gaze and gesture to reference an object (i.e., a wall-mounted display screen) in that space, but the agent's presence was limited to a 2D head and hand on a computer monitor. Similarly, the Rea agent [4] provides a state-of-the-art model of dyadic face-to-face conversation, but bypasses issues of collaboration in dynamic virtual worlds; she can transport herself to and into virtual houses and apartments, and the user can point to some objects within those virtual environments, but the user is not immersed in those environments, and Rea's movement and references within them is very limited. The Cosmo agent [27] includes a sophisticated speech and gesture generation module that chooses appropriate deictic references and gestures to objects in its virtual world based on both spatial considerations and the dialogue context, but the agent and its environment are rendered in 2D and the user does not cohabit the virtual world with Cosmo. The WhizLow pedagogical agent [28] performs tasks in a 3D virtual world, but the agent does not collaborate with students on tasks; the student specifies high-level tasks via menus, and the agent carries them out. Bindiganavale et al. [3] developed a training system that allows multiple virtual humans to collaborate on tasks in a virtual world, but the trainee learns by giving natural language instructions (task knowledge) to the agents and viewing the consequences; she cannot participate in the scenario directly. Each of these systems provides impressive capabilities in its area of research focus, but none of them can interleave task-related behaviors and face-to-face dialogue with humans and virtual humans in dynamic virtual worlds.

To support these capabilities, Steve consists of three main modules: perception, cognition, and motor control [40]. The perception module monitors messages from other software components, identifies relevant events, and maintains a snapshot of the state of the world. It tracks the following information: the simulation state (in terms of objects and their attributes), actions taken by students and other agents, the location of each student and agent, the objects within a student's field of view, and human and agent speech (separate messages indicate the beginning of speech, the end, and a semantic representation of its content). The cognition module, implemented in Soar [24, 34], interprets the input it receives from the perception module, chooses appropriate goals, constructs and executes plans to achieve those goals, and sends motor commands to the motor control module. Steve's cognition module can typically react to new perceptual input in a fraction of a second, so it is very responsive [42]. The cognition module includes a wide variety of domain-independent capabilities, including planning, replanning, and plan execution;

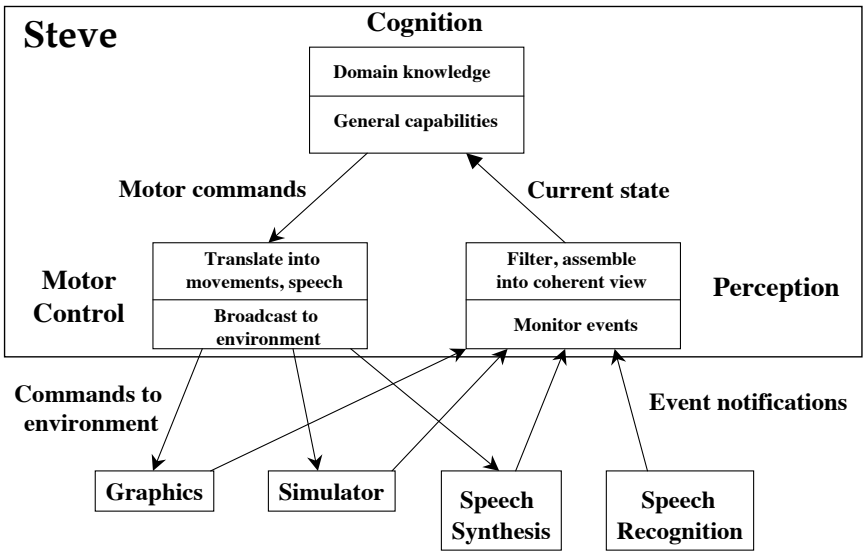

Fig. 2. The architecture of a Steve agent

mixed-initiative dialogue; assessment of student actions; question answering ("What should I do next?" and "Why?"); episodic memory; path planning; communication with teammates [43]; and control of the agent's body. The motor control module accepts the following types of commands: move to an object, point at an object, manipulate an object (about ten types of manipulation are currently supported), look at someone or something, change facial expression, nod or shake the head, and speak. The motor control module decomposes these motor commands into a sequence of lower-level messages that are sent to the other software components (simulator, graphics software, speech synthesizer, and other agents) to realize the desired effects.

Low-level animation of Steve's body runs as software that is linked into the graphics software for the virtual world rather than into Steve. The animation software is controlled by messages it receives from Steve's motor control module. Because the animation software controls the dynamics of all body motions, the motor control module need only specify the type of motion it wants. Steve's animation software is very flexible, placing relatively few constraints on how and when the motor control module can make new demands. This flexibility comes from two properties. First, the animation software generates all movements dynamically; there are no keyframes or canned animations. Second, movements involving different parts of the body can be performed simultaneously, and a new command to a body part interrupts any existing

motion for that part with a smooth transition. The resulting flexibility is an important part of Steve's ability to react instantly to unexpected events.

The cognition module generates Steve's communicative behavior by dynamically selecting its next action from a repertoire of behavioral primitives. To support the needs of task-oriented collaboration, Steve includes the following primitives:

- *Speak:* Steve can produce a verbal utterance directed at a human or another agent. To make it clear to whom the utterance is directed, the motor control module automatically shifts Steve's gaze to the hearer just prior to the utterance. (Task-related events can cause gaze to shift to something else before the utterance is complete.) To make it clear that Steve is speaking, the motor control module automatically maintains a "speaking face" (eyebrows slightly raised and mouth moving) throughout the utterance. Steve has a wide range of utterances, all generated from text templates, ranging from a simple "OK" or "no" to descriptions of domain actions and goals. A message from the speech synthesis software indicates to Steve's perception module when the utterance is complete.
- *Move to an object:* To guide the student to a new object, Steve can plan a shortest path from his current location and move along that path [40]. To guide the student's attention, the motor control module automatically shifts Steve's gaze to his next destination on each leg of the path. In contrast, to simply follow the student around (e.g., when monitoring the student's activities), Steve shrinks and attaches himself to the corner of the student's field of view, so that he can provide visual feedback on their actions.
- *Manipulate an object:* To demonstrate domain task steps, Steve can manipulate objects in a variety of ways. Currently, this includes manipulations that can be done by grasping the object (e.g., moving, pulling, inserting, turning) or using his fingers (e.g., pressing a button, flipping a switch). To guide the student's attention, the motor control module automatically shifts Steve's gaze to the object just prior to the manipulation. State change messages from the simulator indicate to Steve's perception module when the manipulation is complete (e.g., a button's state attribute changing to "depressed").
- *Visually check an object:* Steve can also demonstrate domain task steps that simply require visually checking an object (e.g., checking the oil level on a dipstick or checking whether an indicator light is illuminated). This requires Steve to shift gaze to the object and make a mental note of the relevant property of that object.
- *Point at an object:* To draw a student's attention to an object, or connect a verbal referring expression to the object it denotes, Steve can point at the object. To further guide the student's attention, the motor control module automatically shifts Steve's gaze to the object just prior to pointing at it.

- *Give tutorial feedback:* To provide tutorial feedback on a student's action, Steve indicates a student error by shaking his head as he says "no," and he indicates a correct action by simply looking at the student and nodding. The motivation for shaking the head is to complement and reinforce the verbal evaluation, and the motivation for the head nod is to provide the least obtrusive possible feedback to the student.
- *Offer turn:* Since our goal is to make Steve's demonstrations interactive, we allow students to interrupt with questions ("What next?" and "Why?") or to request to abort the task or finish it themselves. Although they can talk during Steve's utterances or demonstrations, Steve explicitly offers the conversational turn to them after each speech act (which could be several sentences) or performance of a domain action. He does this by shifting his gaze to them and pausing one second. Not only does this give students convenient openings for interruptions, but it also helps to structure Steve's presentations. (Prior to adding this feature, users complained that Steve's presentations were hard to follow because he never paused to take a breath.)
- *Listen to student:* When the student is speaking, Steve can choose to quietly listen. This simply involves shifting gaze to the student to indicate attention.
- *Wait for someone:* When Steve is waiting for someone to take an action (either the student or a teammate in a team training scenario), he can shift gaze to that person (or agent) to indicate his expectation.
- *Acknowledge an utterance:* When a student or teammate says something to Steve, he can choose to explicitly acknowledge his understanding of their utterance by looking at them and nodding. The speech recognizer does not provide recognition of intermediate clauses, so Steve is limited to acknowledging understanding of entire utterances.
- *Drop hands:* When Steve is not using his arms and hands, he can drop them back down to hang loosely at his sides. Although there is evidence that such a move can convey a conversational signal (i.e., end of turn) [12], Steve does not currently use this behavior for that purpose; it simply means he has nothing else to do with his hands (such as pointing or manipulating).
- *Attend to action:* When someone other than Steve manipulates an object in his environment, Steve automatically shifts his gaze to the object to indicate his awareness. Unlike all the above behaviors, which are chosen deliberately by the cognition module, this behavior is a sort of knee-jerk reaction invoked directly by the perception module. Because an object manipulation is a very transient event, our design rationale was to react as quickly as possible.

Steve's main challenge is generating coherent behavior, and the key to addressing this challenge is to maintain a rich representation of context. The ability to react to unexpected events and handle interruptions is crucial for task-oriented collaboration in dynamic virtual worlds, yet it threatens the

coherence of an agent's behavior. A good representation of context allow an agent to be responsive while maintaining its overall focus. Steve maintains two separate but complementary types of context:

- *Task Context:* Steve models tasks using a hierarchical, partial-order plan representation. As a task proceeds, Steve continually monitors the state of the virtual world, and it uses the task model to maintain a plan for how to complete the task, using a variant of partial-order planning techniques [40]. This allows Steve to revise its plans to adapt to unexpected events.
- *Dialogue Context:* The dialogue context represents the state of the interaction between a student and Steve, including whether the student and/or Steve is currently speaking; a focus stack [20] representing the hierarchy of tasks, subtasks, and actions in which the student and Steve are currently engaged; the state of their interaction on the current task step (e.g., the state might be that Steve has explained what must be done next but neither he nor the student has done it); a record of whether Steve or the student is currently responsible for completing the task (this task initiative can change during a mixed-initiative interaction); the last answer Steve gave, in case the student asks a follow-up question; any pending obligations (i.e., student requests or questions that Steve has not yet addressed); and the actions that Steve and the student have already taken [42].

Based on the current task and dialogue contexts, Steve can choose his next action to fill one of three roles. First, he can respond to the student. This includes responding to a student's request, giving them tutorial feedback on their action, or simply listening when they are talking. Second, Steve can choose for himself how to advance the collaborative dialogue. This includes things like suggesting the next task step, describing it, or demonstrating it in cases where the student did not explicitly request such help. Third, Steve can choose a turn-taking or grounding act [50] that helps regulate the dialogue between the student and himself without advancing the task. This includes offering the student the conversational turn or acknowledging understanding of an utterance with a head nod.

Several such actions may be appropriate at any given moment, so priorities allow Steve to choose the most appropriate. The highest priority is to respond to the student. If no such actions are proposed, the next priority is to perform any relevant conversational regulation action. However, if an opportunity for a conversational regulation action is missed due to a higher priority action for responding to the student, it will not be deferred and performed later. Only when neither of these types of actions is proposed will Steve take the initiative to advance the task collaboration, and he only does that when he has the task initiative. Traum proposed a similar priority scheme in his model of spoken task-oriented dialogue [49].

The original version of Steve, described in this section, did not include a model of emotions, and expressive behavior was not a primary research focus. However, Steve's broad capabilities have served as a valuable founda-

tion for our current work on virtual humans, described later in this chapter. The expressive behaviors that we have integrated into Steve are derived from two separate research projects: Émile (Jack and Steve) and Carmen's Bright IDEAS.

3 Jack and Steve

Jack and Steve was an exploration in the extent to which plans and plan reasoning could inform rich models of expressive social behavior. The Jack and Steve system was predicated on two basic claims: 1) that plans and plan reasoning can mediate expressive behavior and 2) that small biases in how plans are evaluated and generated can result in large systematic differences in agent behavior.

With these goals in mind, Jack and Steve differed in several respects from the preceding Steve system. To support biases in plan evaluation, Jack and Steve incorporated a richer plan representation, including decision-theoretic information to inform the evaluation process and meta-planning capabilities. To translate small biases into large external variations, Jack and Steve focused on plan generation, whereby these biases could be magnified over the multiple steps involved in the generation process. In contrast, Steve focused more on plan execution and repair, so its plan generation algorithm was less general. The systems also differed significantly in terms of their inter-agent behavior. Steve focused on collaborative interactions between agents and human users, whereas Jack and Steve explored how biases in the plan generation process could support a variety of non-collaborative interactions as well, but, as the focus was exploring systematic differences in joint behavior, agents only interact with other computational agents and not with human users.

The Jack and Steve system led to two key innovations that have influenced our subsequent agent designs: Émile [16], a plan-based model of emotional appraisal, and *socially situated planning*[17], whereby an agent may alter its goal-directed behavior based on features of the social context. These models were integrated with the animation system developed for the Steve system, described above, augmenting them with a limited ability to generate facial expressions, expressive gestures, and emotionally biased speech synthesis.

Jack and Steve was motivated by a convergence of several distinct bodies of research. First, psychological theories of emotion emphasize the pervasive role of emotions in social interactions and suggested that human emotions are mediated by some form of goal-directed reasoning. Cognitive appraisal theory, in particular, argues that emotions arise from an assessment process that characterizes how events impact goals along several abstract dimensions such as goal-relevance, goal-congruence, and likelihood [45]. Second, psychological theories of personality illustrate that people of different personality types will appraise and respond to events quite differently, and that goal-directed reasoning may mediate this process as well. These theories argue that relatively

small biases in appraisal or response might lead to large systematic differences in outward behavior. For example, conscientious individuals tend to accept greater personal responsibility for joint goals, which in turn can lead to the construction of quite different plans and collaborative interactions than might result from a non-conscientious individual [38]. Finally, artificial intelligence theories of collaborative behavior have begun to build formal models of social interaction in terms of goal-directed reasoning. Work on shared plans [19] and joint intentions [10] illustrate how reasoning about interactions between plans, and representing beliefs, obligations and commitments, could motivate social behavior. Jack and Steve joined these strands of research into a system that utilized plans and plan reasoning to model an agent's relationship to its physical and social environment, and to support systematic individual variations in how this relationship was conceived.

3.1 Motivating Example

The Jack and Steve application domain centered on the antics of two Southern Californian roommates, Jack and Steve. The agents engaged in unscripted interactions via simulated speech, and a user could explore a variety of interactions by altering internal characteristics of either agent. For example, a user could alter an agent's goals (e.g, to have fun or to make money) and alter characteristics of their personality (e.g., are they cooperative or rude).

In this motivating example, Jack's goal is to make money, he views Steve as a friend, and treats him fairly. Steve wants to surf, views Jack as a friend, but tends to be rude in his dealings. All of these terms have a specific technical definition discussed below. Both agents develop different plans but have to contend with a shared resource. Besides performing task level actions, the agents engage in speech acts and generate gestures, facial expressions, and affective speech modulation based on properties of the social context.

What follows are annotated traces of two separate runs of the system where the only difference is a change in the personality of the Steve agent. In the first trace he treats Jack rudely and in the second he treats him fairly. The agents generate speech via simple template filling and agents actually communicate with each other through a stylized plan-communication language. Figure 3 illustrates the mental state of each agent at some point in the interaction. White boxes indicate individual actions and arrows indicate the establishment of an action's preconditions by another action's effects, or a threat to some precondition's establishment by an intervening action that negates the establishing effect. Blue boxes indicate "plans," which are sets of actions treated as a conceptual unit by the social layer. The emotion windows illustrate each agent's current emotional state, characterized in terms of intensity values along a set of basic emotions.

Rude Interaction:

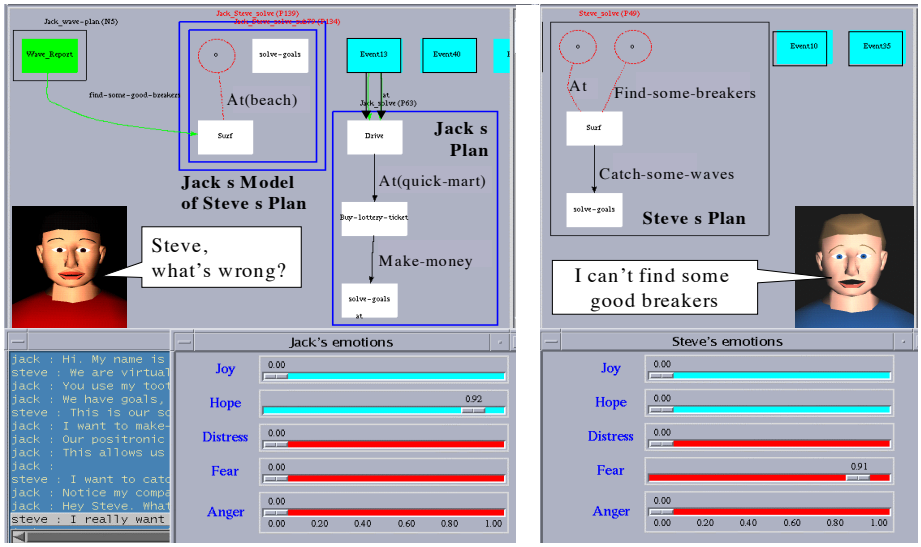

Fig. 3. Representation of Jack and Steve's appraised state, including each agent's base-level task network, plans, and emotional state.

Jack: I want to make-some-big-money. *[Begins generating a plan for this goal. Displays a concerned expression, scratches his head, then, after devising a complete plan, displays a hopeful expression.]*

Steve: I want to catch-some-waves. *[Begins generating a plan for this goal. Looks concerned, scratches head, and continues to look concerned. Surfing is important to Steve and he cannot devise a satisficing plan.]*

Jack: *[Perceives Steve's display of concern and generates an information request.]* Hey Steve, what's wrong?

Steve: *[Identifies the feature in plan memory contributing to the most intense negative emotional excitation. Communicates the associated plan in a distressed tone of voice.]* I want to catch some waves but can't find any good breakers.

Jack: *[Incorporates Steve's plan into plan memory and locates relevant information. Jack knows of an event that establishes Steve's blocked subgoal.]* Steve, does it help that someone did say there's some great waves near the pier?

Steve: *[Incorporates the communicated event into plan memory. Completes a plan to go surfing and looks hopeful.]*

Jack: *[Perceives Steve's change in expression and seeks to confirm his expectation that the information he provided helped Steve.]* So that information helped?

Steve: *[Handles Jack's information request.]* Yes Jack. I plan to drive the car to the beach, then I plan to surf-my-brains-out.

Jack: *[Incorporates Steve's revised plan and finds a conflict with his own plans. Based on personality, Jack attempts to negotiate a fair solution.]* Wait a second. Our plans conflict. I plan to drive the car to the-quicky-mart then I plan to buy a-lottery-ticket.

Steve: *[Incorporates Jack's plan and recognizes the same interaction. Based on personality model, Steve responds to the interaction differently. He devises a plan that satisfies his own goals without regard to any conflicts it may introduce in Jack's plans. Steve exits stage right.]* Later dude, I'm driving the car to the beach.

Jack: *[Perceives that the car has departed without him. Looks angry. Says in angry voice:]* I want to kill-my-roommate.

Cooperative Interaction:

In this second interaction, the user replays the interaction but first alters Steve's personality to treat Jack fairly. The agents have identical goals and knowledge, but due to differences in their social appraisals the interaction is quite different. The interaction is identical up to the point that Jack detects an interaction between the plans:

Jack: *[Incorporates Steve's revised plan and finds a conflict with his own plans. Based on personality, Jack attempts to negotiate a fair solution.]* Wait a second. Our plans conflict. I plan to drive the car to the-quicky-mart then I plan to buy a-lottery-ticket.

Steve: *[Incorporates Jack's plan and recognizes the same interaction. Based on Steve having somewhat lower social status, he takes the initiative in repairing the conflict.]* Well, I could change my plans. *[Looks concerned, scratches head, then devises a possible joint plan.]* I have a suggestion. Could you drive the car to the-quicky-mart with-me then I could drive the car to the beach.

Jack: *[Incorporates Steve's suggested joint plan, determines that it is consistent with his own plans, and agrees to form a joint commitment to the shared plan.]* Sounds good to me.

3.2 Plan-based Social Appraisal

As discussed earlier, The Jack and Steve system was predicated on two basic claims: 1) that plans and plan reasoning can mediate expressive behavior and 2) that small biases in how plans are appraised and generated can result in large systematic differences in agent behavior. Jack and Steve supported such expressive and flexible interactions by implementing social reasoning as a layer atop a general-purpose partial-order planning system [1, 51]. The planning system provides domain-independent representations of world actions in terms of preconditions and effects, and provides general reasoning mechanisms that construct partial plans, repair interactions between them, and oversee plan execution. The social layer manages communication and biases plan generation and execution in accordance with the social context (as assessed within this social layer). In this sense, social reasoning is formalized as a form of meta-reasoning.

To support a variety of social interactions, the social reasoning layer must provide a rich model of the social context. The social situation is described in terms of a number of static and dynamic features from a particular agent's perspective. Static features include innate properties of the character being modeled (social role and a small set of personality variables). Dynamic features are derived from a set of domain-independent inference procedures that operate on the current mental state of the agent. These include the set of current communicative obligations, a variety of relations between the plans in memory (e.g., your plans threaten my plans), and a model of the emotional state of the agent (important for its communicative role).

One novel aspect of the system is the way in which the social layer fundamentally alters the planning process. Grosz and Kraus [19] show how meta-level constructs like social commitments can act as constraints that limit the planning process in support of collaboration (for example, by preventing a planner from unilaterally altering an agreed-upon joint plan). Jack and Steve went beyond this to show how to model a variety of "social stances" one can take towards other individuals based on one's role in an organization and other dispositional factors. In terms of planning, rather than simply being cooperative, the social layer can bias planning to be more or less considerate to the goals of other participants. In terms of communication, agents can vary in terms of how much initiative or control they can take over the interaction, from bossy agents that try to tell others what to do to more passive agents that meekly avoid interactions or social conflicts.

3.3 Social Context

As in the preceding Steve system, Jack and Steve maintains a rich representation of the social context to drive coherent behavior. Domain independent *appraisal rules* map features of an agent's current plan knowledge into a current *social context*. Besides static features of the social context set by the user, such as the agents' goals and personality, the social context can be divided into the following distinct components.

Plan Context: The plan context plays an analogous role to the Task Context in the original Steve system. The plan context represents information about the plans agents are entertaining, as well as meta-level information about the status of these plans. The Jack and Steve system incorporated a different plan representation than the original Steve system. While both systems used general plan representations, Jack and Steve also incorporated a decision-theoretic model, representing the likelihood and utility of various plans (which is quite useful in modeling the intensity of emotional responses). Unlike Steve's plans, which are hierarchical, Jack and Steve adopted a simpler non-hierarchical plan representation but included an explicit model of meta-plan reasoning.

Jack and Steve's base-level planning layer represents future-directed actions that an agent is aware of (whether they come from its own planning

or are communicated from outside) as a single plan in the classical planning sense (i.e., a partially-ordered set of actions with establishment and threat relations between actions), henceforth referred to as the *task network*. This allows the base-level planner to reason about the interrelationship between these activities. However, at the social level, subsets of this task network are explicitly treated as distinct plans in the commonsense use of the term (i.e, a coherent set of actions directed towards a goal), henceforth referred to as *plans*. Plans at this social layer may belong to the agent or may corresponding to (what the agent believes to be) plans of other gents.

The plan context also represents a number of meta-level relations between these plans. Plans can contain threats if the actions within a plan threaten each other and the plans of one agent can introduce threats or be threatened by the plans of another agent (such relations are computed using the basic plan-evaluation routines provided by standard planning systems). Plans of one agent are deemed *relevant* to the plans of other agents if they may causally interact [11].

Emotional Context: Unlike the preceding Steve system, the Jack and Steve system maintains a representation of the agents' emotional state. The model of emotional reasoning that supports this, Émile, has been described extensively elsewhere [16, 18]. Émile adopts the cognitive view of emotions as a form of plan evaluation, relating events to an agent's current goals (c.f., [35, 25]). As in other appraisal-based computational models of emotion [13, 33], Émile classifies events in terms of a set of *appraisal variables*:

- goal relevance - are the consequences of an event relevant to an organism's goals
- desirability - how desirable are the consequences
- likelihood - how likely are the consequences
- causal attribution - who is the causal agent underlying the event and do they deserve credit or blame

Unlike prior computational models, Émile reified these variables in terms of domain-independent features of an agent's plans in memory. Émile contains a set of recognition rules that scan an agent's internal representations and generate an appraisal frame whenever certain features are recognized. For example, when Steve states he will drive the car to the beach, the effect of this potential action (that the car is no longer at home) threatens Jack's plan to get to the quicky mart. The existence of a threat to an important goal is interpreted as an undesirable event, which ultimately gets mapped into a fear response based on the likelihood of the threat and the importance of the goal.

Many appraisals may arise from the current plan context (e.g., Steve may be simultaneously hopeful that he will surf but fearful that Jack may abscond with the car). Individual appraisals are collected together by class and their intensities summed into an overall emotional context. Different subsets of appraisals can be aggregated and associated with meta-level constructs, allowing Émile to compute an agent's overall state, track the emotions arising from a

specific plan, or make estimates of the overall emotional state of other agents (given an understanding of their goals and plans). Each of these aggregate states is represented as a real-valued vector representing the intensities of different emotional states (Fear, Joy, etc.) and Émile dynamically modifies this state as appraisals change in response to the current world situation and the state of plans in memory.

Communicative Context: The communicative context tracks what information has been communicated to different agents and maintains any communicative obligations that arise from speech acts. When one agent communicates a plan to another, the social layer inserts the plan into the task network and records that the recipient knows this plan. This belief persists until the sender's planning layer modifies the plan, at which point the social layer records that the recipient's knowledge is out of date. If one agent requests another's current plans, the social layer represents a communicative obligation: the fact that the recipient of the request owes a response is recorded in each agent's social layer (though whether the recipient satisfies this obligation is up to its own social control program).

The communicative context roughly corresponds to the earlier Steve system's notion of dialogue context in the sense that both are concerned with representing a history of preceding communication and its impact on current beliefs and pending obligations. The Steve system, however, focused more on dialogue related to hierarchical task execution (e.g., the focus stack), whereas Jack and Steve focus more on the relationship of dialogue acts to plan generation.

3.4 Social Operators and Social Stances

Social operators are actions that occur at the social level. These are subdivided into meta-planning operators and communicative operators. By triggering these operators based on features of an agent's social context, the Jack and Steve system could represent a number of distinct social behaviors.

Meta-planning operators alter the way the planner operates at the base-level. Meta-planning operators allow the social level to react and manipulate plan objects, populate them with subsets of the task network, and alter how the planner operates on those subsets. For example, if Steve communicates his plan to Jack, Jack could create a new plan object ("Steve's Plan") and populate it with the set of communicated actions. Different plans could be treated differently by allowing or disallowing certain types of planning modifications. Classical planning algorithms can be viewed as a sequential decision process: some critiquing routines identify a set of problems with the current plan network and propose a set of modifications that resolve at least one of these problems (an action should be added, these actions should be reordered, etc.); one modification is applied and the process continues [23]. The social level might disallow any changes to a plan (corresponding to the idea that

the agent is committed to the plan), or it may allow more subtle variations in how plans are changed by constraining the set of allowable modifications.

Communicative operators correspond to a set of speech acts that an agent may use to communicate with other agents. As they are defined at the meta-level, they can operate on plans only as an atomic structure and cannot make reference to components of a plan (although one has the option of breaking a plan into explicit sub-plans). Some speech acts serve to communicate plans (one can INFORM another agent of one's plans, REQUEST that they accept some plan of activity, etc.). Other speech acts serve to change the state of some previously communicated plan (one can state that some plan is under revision, that a plan is acceptable, that it should be forgotten, etc.). Communicative primitives also include non-verbal communication, such as gestures.

From the standpoint of modeling expressive behavior, the most novel contribution of the Jack and Steve system was the way social operators could alter the way the planner handles interactions between plans of different agents, thereby implementing the idea of a social stance. A number of distinct social stances could be modeled and were organized along four roughly orthogonal dimensions: *conscientiousness, dominance, sociability,* and *independence.* These stances are all implemented as search control strategies, limiting certain of a planner's threat resolution options or the agent's communication options.

Conscientiousness impacts the extent that an agent respects the goals and plans of other agents. A non-conscientious (rude) agent only considers threats to his own plans and discounts any threats that his own actions introduce into the plans of other agents. For example, the rude Steve agent runs to grab the keys before Jack gets a chance to take the car. This corresponds to the threat resolution strategy of promotion, whereby a threatened action is moved before the threat. A conscientious agent wouldn't consider *promotion* as it prevents Jack's plans from succeeding (in planning terminology this is an instance of *brother-clobbers-brother-goal*); however a rude agent would discount this other-directed threat.

Dominance impacts whether an agent is willing to dictate actions to other agents. For example, a dominant agent would freely introduce actions into the plans of other agents, or incorporate steps into his own plans that other agents are expected to perform. In contrast, a meek agent would avoid these options and tend to work around interactions. For example, a meek Steve might find some other way to get to the beach or simply stay home.

Sociability relates to how readily an agent communicates to resolve conflicts or to provide potentially useful information. Social agents communicate whenever they encounter interactions between plans while asocial agents would try to resolve conflicts without communication. For example, an asocial agent could resolve the resource conflict involving the car by simply taking the car to the quicky mart before the other agent gets a chance to take it to the beach.

Finally, agents vary in terms of their independence. An independent agent would refuse to develop plans that depended on the actions of other agents.

For example, the plan of riding to the quicky mart together would be ruled out by an independent agent.

Beyond social stances, meta-operators allow the social level to create and manipulate plan objects. Plans can be created and destroyed, and they can be populated with new goals and with activities communicated by other agents. Another set of meta-operators determines whether the planning algorithm can modify the activities in one of these plan objects. One can make a plan modifiable, allowing the planner to fix any flaws with that plan, or one can freeze its current state (as when adopting a commitment to a certain course of action. One can also modify the execution status of the plan, enabling or disabling the execution of actions within it.

Distinct personalities and social stances are implemented via a set of *social rules* that execute sequences of social operators based on appraised features of the social context. These rules can be viewed as a simple social domain theory or, alternatively, as a simple reactive plan.

The Jack and Steve system included about thirty social rules. A few examples are listed here.

Social-Rule: plan-for-goal
IF I have a top-level ?goal
THEN
  Do-Gesture(Thinking)
  Say(to-self, "I want to ?goal")
  ?plan = create-new-plan(?goal)
  enable-modification(?plan)

The *plan-for-goal* rule creates a new plan object for an agent's top-level consisting of one dummy step that has the goal as its precondition, and, by enabling modification, allows the base-level planner to add actions to the plan in order to achieve the goal. The rule also triggers an utterance ("I want to ...") and an expressive "Thinking" gesture (implemented by a motor procedure that turns the agent's head up and to the side and raises one arm to scratch the head).

Social-Rule: commit-to-plan
IF I have ?plan
  AND I am currently modifying ?plan
  AND the ?plan is free of threats
THEN
  Do-Gesture(Nod-Head)
  commit-to(?plan)
  disable-modification(?plan)

If at some point the base-level planner successfully constructs a threat-free plan to achieve to goal, the *commit-to-plan* rule commits to the plan and prevents the planner from making further modifications. Note that the definition

of "threat-free" is dependent on the agent's social stance (for example, a rude agent may perceive his plan as threat-free even though it is clobbering steps of another agent's plans).

Social-Rule: help-friend
IF I have a ?plan that is relevant-to the plan of another ?agent
  AND I am friends with ?agent
  AND I am socially-adept
  AND the ?plan is not known to ?agent
THEN
  Do-Gesture(Look-at ?agent)
  SpeechAct(INFORM, ?plan ?agent)

If an agent has been defined to be socially-adept and they are aware of some information that may causally impact the plans of another agent, the *help-friend* rule ensures that this information is communicated to the other agent.

Social-Rule: you-cause-problems-for-me
IF I have a ?plan
  AND ?you have ?your-plan
  AND my ?plan is threatened by ?your-plan
  AND I am committed to my ?plan
  AND I am not meek
  AND ?you don't know my ?plan
THEN
  Say(?you, "Wait a second, our plans conflict")
  SpeechAct(INFORM, ?plan, ?you)

If an agent has committed to a plan and discovers that the actions of another agent are threatening the plan's execution, a non-meek agent should communicate his own plans with the unstated expectation that the other agent will respond cooperatively.

Many of these rules, such as help-friend, correspond to standard conventions in collaborative planning. What is novel about Jack and Steve, however, is the idea of differentially applying them depending on features of an agent's personality, allowing, for example, a flexible gradation between cooperative and non-cooperative behavior. Collectively, these rules form a sort of social domain theory and, by explicitly representing social context and social operators, Jack and Steve facilitates the easy construction of different mappings between them and thus easy experimentation with different social theories.

3.5 Bodily Expression

The Jack and Steve system's chief contributions are in its internal process models of how emotion gets appraised and planning gets altered by social

stances. However, this internal machinery must be manifested externally to have an impact on the user. Jack and Steve's reasoning mechanism was connected to the Steve agent body described earlier. This provided non-photorealistic 3D human-like bodies that included control of body movements (including procedural control of gaze, pointing and grasping) and procedural control of facial expressions (including control of eyebrows, eyelids and mouth characteristics). It also included a text-to-speech system with some coarse control over the characteristics of speech that could give some sense of emotional speech. Through these controls we developed a small repertoire of exaggerated facial poses to convey basic emotions and a set of arm and facial gestures to indicate other mental processes. This included pointing to the air and nodding when successfully completing a plan, scratching one's head when developing a plan, and winking when irrecoverably destroying another agent's plans.

Jack and Steve were never formally evaluated but anecdotal evidence suggests that people could easily recognize certain basic differences in personality (e.g. rude versus cooperative) through the different trajectory of the interactions and the type of plans agents developed, although not all combinations of traits led to recognizable differences. Facial expressions and gestures seemed primarily useful for conveying information about the agent's appraised internal state and apparently added to the perceived humorousness of the interaction.

We subsequently implemented a version of the system using more photorealistic faces which people, interestingly, found rather less funny and more disturbing. This was likely due to the crude control we had over facial expressions. In both systems, characters would hold a fixed facial expression for several seconds. In the non-photorealistic Steve graphical bodies, people seemed to find this acceptable. However with the more photorealistic faces people felt the characters were "creepy" or "maniacal." This reinforces the conventional wisdom that the drive toward photorealism in graphical models will demand considerably more attention to the form and dynamics of physical expressions.

4 Carmen's Bright IDEAS

Carmen's Bright IDEAS (CBI) was an agent-based system designed to realize an Interactive Pedagogical Drama (IPD) [31], an approach to learning that immerses the learner in an engaging, evocative story where she interacts openly with realistic characters. The pedagogical goal of CBI was to help mothers of pediatric cancer patients deal with the many stresses they face due to their child's illness. A mother learns by making decisions or taking actions on behalf of a character in the story, and sees the consequences of her decisions subsequently played out. To bring this pedagogy to life, the drama mirrors the mother's own problems. In the CBI story, the various stresses Carmen is facing are revealed, including her son's cancer, her daughter Diana's temper

tantrums, work problems, etc. The drama foregrounds these stresses and allows the learner to interactively influence how Carmen copes with them. To facilitate open interaction, the characters in CBI were realized as autonomous agents.

Many of the differences with the previously discussed systems stem from the fact that the agents in CBI realize a social drama about very stressful issues. In particular, CBI's drama explores how the main character agent develops cognitively and emotionally. In contrast, the focus for the previously discussed Steve agent was on performing a procedural task in an immersive environment. Therefore drama and character development was not a central concern.

As a consequence of this dramatic goal, the design of the agent models in CBI was rooted in psychological research in stress and causes of emotions. In particular, like Émile, cognitive appraisal theory influenced the design of the CBI agents, though the two systems realize different, complementary aspects of appraisal theory. The focus in Émile was on the task-oriented causes of emotion. In CBI, there was a more pressing requirement to model the causes of emotion that stem from what psychologists call an ego identity, an individual's concerns for loved ones, for how others perceive them, for performing their social roles well, and for measuring up to their personal ideals. Emotions stem from how events impact these concerns. Further, CBI agents needed to also model the consequences of emotions – how people cope with difficult emotional stresses in both adaptive and maladaptive ways as well as how to learn better ways of coping.

Also, the animated agents needed effective ways to convey the impact of emotion on both the agent's dialog and physical behavior. In particular, this required developing models of how complex, sophisticated emotional stress processes are revealed over time in coping behavior and dialog. Although other systems have addressed expressive behavior, this modeling of human coping and its dynamic impact on behavior set CBI apart. The concern for expressive behavior that reveals underlying dynamics grew out of the fact that CBI was being designed for a clinical trial with mothers of pediatric cancer patients. The dynamics would potentially benefit believability of the agents and facilitate the learner's identification with the agents. In addition, it might further the learner's understanding of the underlying emotional and coping processes that the agents were modeling, which in fact was part of the pedagogy.

4.1 The drama of Carmen's Bright IDEAS

CBI is a three act interactive drama. In the key act, Carmen discusses her problems with a clinical counselor, Gina, who suggests she use a problem solving technique called Bright IDEAS to help her find solutions. Note each letter of IDEAS refers to a separate step in the problem solving method: Identify a solvable problem, Develop possible solutions, Evaluate your options,

Act on your plan and See if it worked. Bright refers to the need for a positive attitude. Figure 4 is a shot of Carmen and Gina in Gina's office. With Gina's help, Carmen goes through the initial steps of Bright IDEAS, applying the steps to one of her problems, and then completes the remaining steps on her own. The final act reveals the outcomes of Carmen's application of Bright IDEAS.

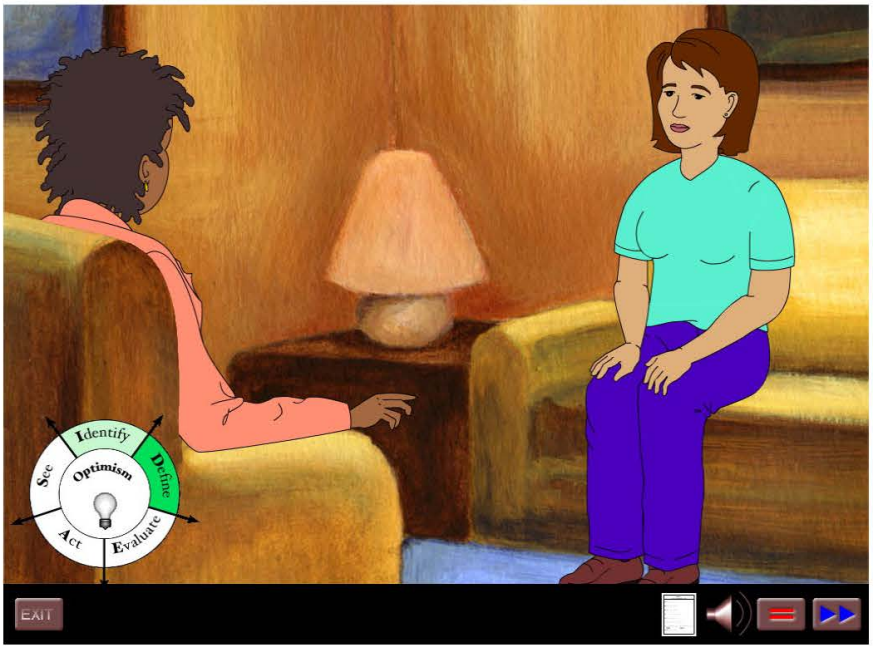

Fig. 4. Carmen (right) speaking with Gina (left).

Central to the drama's tension is the interaction between Gina, Carmen and the learner. The interaction model we designed for CBI is what we call a rubber-band model. See Figure 5. Both Gina and the learner exert influence over Carmen but the influence is partial and mediated by Carmen's own cognitive and emotional dynamics. Thus we characterize this influence as rubber-bands. It is Gina's job to keep the social problem solving on track so that the story proceeds to a successful outcome by effectively responding to Carmen's cognitive and emotional state, at times motivating her through dialog to work through the steps of IDEAS on some problem or alternatively calming or reassuring her. The human mother interacts with the drama by making choices for Carmen such as what problem to work on and how she should cope with the stresses she is facing. The learner can choose alternative internal thoughts for Carmen, such as "I hope this helps with Diana."

These are presented as thought balloons (see Figure 6). Both Gina's dialog moves and the learner's choices influence the cognitive and emotional state of the agent playing Carmen, which in turn impacts her behavior and dialog, perhaps in conflicting ways. The cognitive and emotional dynamics within the Carmen agent ensures that Carmen's behavior is believable at all times, regardless of how Gina and the learner may be influencing her.

In this interaction model, the Gina agent is both an on-screen character and the drama's director. The social interaction between agents is driven by the Gina agent's persistent goal to motivate the Carmen agent. Therefore, Gina typically takes the initiative. If the Carmen agent is distressed, she requires considerable prompting, praise and guidance from Gina. But as she is reassured about IDEAS, she will begin to "feel" hopeful and may engage the problem solving without explicit prompting.

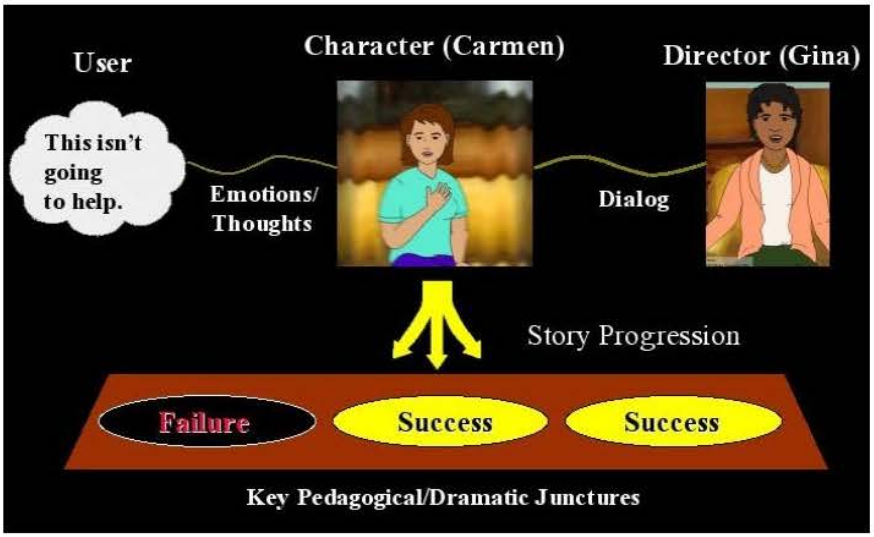

Fig. 5. Rubber band interaction model.

The combination of Gina's motivation of Carmen through dialog and the learner's emotional impact on Carmen creates tension, a rubber-band tug-of-war between Gina's attempts to motivate Carmen and the initial, possibly less positive, attitudes of the Carmen/learner pair. As the learner plays a role in determining Carmen's attitudes, she assumes a relationship in this tug-of-war, including, ideally, an identification with Carmen and her difficulties, a responsibility for the onscreen action and perhaps empathy for Gina. If Gina gets Carmen to actively engage in applying the IDEAS technique with a positive attitude, then she potentially wins over the learner, giving her a

positive attitude. Regardless, the learner gets a vivid demonstration of how to apply the technique. The design also allows the learner to adopt different relationships to Carmen and the story. The learner may have Carmen feel as she would, act the way she would or "act out" in ways she would not in front of her real-world counselor.

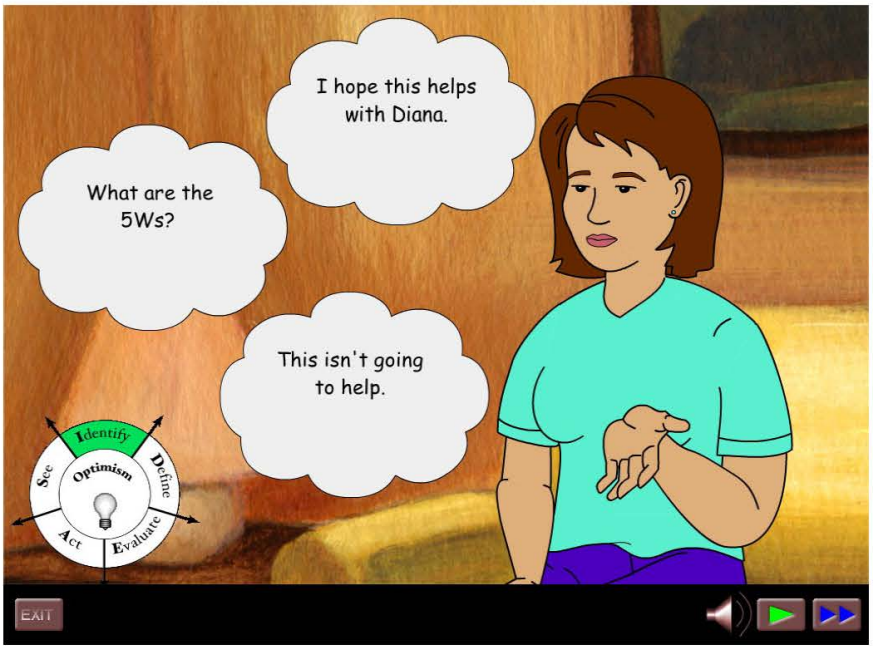

Fig. 6. Learner influences Carmen by selecting Thought Balloons.

4.2 CBI Agent Models

Technically, the basic design for interactive pedagogical drama includes five main components: a cast of autonomous character agents, the 2D or 3D puppets which are the physical manifestations of those agents, a director agent, a cinematographer agent, and finally the learner/user who impacts the behavior of the characters. Animated agents in the drama choose their actions autonomously but also follow directions from the learner and/or a director agent. Director and cinematographer agents manage the interactive drama's onscreen action and its presentation, respectively, so as to maintain story structure, achieve pedagogical goals, and present the dynamic story so as to achieve best dramatic effect. Here, the discussion will focus on the onscreen

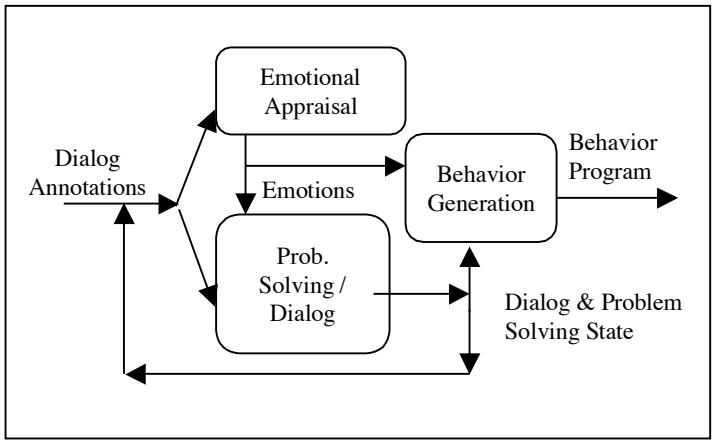

Fig. 7. Agent Architecture.

character agents. In CBI, one of the onscreen character agents, Gina, also serves as director. Further discussion of the cinematographer agent can be found in [31].

Each onscreen character is realized by an agent architecture that has modules for problem solving, dialog, emotional appraisal and behavior generation. See Figure 7. The problem solving module is the agent's cognitive layer, specifically its goals, planning and deliberative reaction to world events. The dialog module models how to use dialog to achieve goals. The emotional appraisal module determines how the agent emotionally evaluates (appraises) events. Finally, behavior generation constructs the agent's behavior and passes that behavior on to an animation program (not shown). Each of these modules will be discussed in greater detail in subsequent sections.

There are several novel pathways in the architecture worth noting. It is possible for the agent to say something and emotionally and cognitively react to the fact that it has said it, since the agent's own dialog feeds back as input. Emotions impact problem solving, dialog and behavior. Finally, there are multiple inputs to behavior generation, from emotional appraisal, dialog and problem solving, all competing for the agent's physical resources (arms, legs, mouth, head, etc.). For instance, the dialog module derives dialog that it intends to communicate, which may include an intent to project an associated emotion. This communication may be suggestive of certain nonverbal behavior for the agent's face, arms, hands etc. However, the agent's emotional state

derived from emotional appraisal may suggest quite different behaviors. As we will discuss, behavior generation mediates this contention.

An example demonstrates how these pathways work within the architecture. At one point, Gina asks Carmen why her daughter is having tantrums. Carmen tends to feel anxious about being judged a bad mother and the learner may choose a thought that reinforces this anxiety. Carmen copes (problem solving) by dismissing the significance of the tantrums (dialog model): "She is just being babyish, she wants attention." Based on Carmen's dialog and emotional state, behavior generation selects relevant behaviors (e.g., fidgeting with her hands). Her dialog also feeds back to emotional appraisal. She may now feel guilty for "de-humanizing" her child, may physically display that feeling (behavior generation) and then go on to openly blame herself. Carmen can go through this sequence of interactions solely based on the flux in her emotional reaction to her own behavior. Gina, meanwhile, will emotionally appraise Carmen's seeming callousness and briefly reveal shock (e.g., by raised eyebrows), but that behavior is quickly overridden if her dialog model decides to project sympathy.

As noted in Figure 7, the agents use dialog annotations to communicate. In order to maximize expressive effect of such dialog, recorded dialog of voice actors was used instead of speech synthesis. A significant amount of variability in the generated dialog is supported by breaking the recordings into meaningful individual phrases and fragments and by recording multiple variations (in content and emotional expression). There are 480 dialog fragments in the clinical trial version of CBI. The agents compose their dialog on the fly, using annotations attached to the fragments to understand each other and decide how to respond. The agents experience each fragment's annotation in order, so their internal state and appearance can be in flux over the dialog segment.

Emotional appraisal and dialog are depicted in Figure 7 as happening in parallel. However the phrase-by-phrase dialog generation and comprehension of the associated annotation tags in practice results in appraisals being interleaved with the dialog. As this sequential process unfolds phrase by phrase, a behavior program is being incrementally constructed by the Behavior Generation. It is useful to understand this sequencing because it helps determine how emotion relates to other components of the architecture and therefore how the subtlety and dynamics of emotional expression in CBI is realized. In particular, the sequencing is closely related to the resulting expressive behavior program that will be discussed in the Behavior Generation section.

Agents process dialog in the following order:

1. "Hear" Dialog Line (other speaker) if there is one.
2. Appraise dialog from step 1 and pass result to Behavior Generation.
3. Decision-making form intent to perform a dialog act based on some list of phrases to speak.
4. If agent wants to speak goto 5 otherwise goto 1.
5. Appraise step 3 decision-making and pass result to Behavior Generation.

6. Compute dialog pause based on emotional state and pass to Behavior Generation.
7. Pass phrase and annotations to Behavior Generation to build the part of the behavior program that will be executed in parallel with this phrase.
8. Appraise phrase just spoken and pass to Behavior Generation.
9. If phrases remain go to 6 else inform Behavior Generation that dialog turn is over and goto 1.

The Appraisal at (2) is the starting emotional appraisal. It sets emotional state and consequently informs behavior generation, and thus indirectly manages the initial face, head turns and body expression (posture) of the agent, as well as impacting the decision making that precedes its own dialog. Decision making at (3) comes up with a plan to say something. The appraisal at (5) sets emotional state and expressions based on decision making at 3 - the dialog intent/content of the entire response. This may in turn update face expression, head position and body focus. The pause at (6) simply manipulates the time between the agent's phrases based on emotional state, allowing expressive pauses in the dialog.

The part of the behavior program built at (7) creates a parallel and sequential structure that includes gestures, facial expressions, head movements, blink patterns, posture shifts and speaking of the appropriate surface phrase. Because of animation infrastructure, the behavior program is incrementally built by these steps but execution only happens when the complete program is constructed for the dialog turn. The appraisal at (8) sets emotional state in reaction to content relayed by each phrase.

The dialog annotations are not designed to fully describe the dialog but rather constitute an abstract level at which the agents reason about how to react to each other's as well as their own dialog. Annotations include:

- Dialog content: Dialog Act, Speaker and Addressee
- Emotional content: Coping Act (e.g. denial)
- Propositional content: Main referent (e.g. Diana) and Topic (e.g., temper tantrums)
- Performance content: Referential structure (e.g., "me" indicates speaker is referring to self as "I feel...", "me other" indicates speaker is referring to self and someone else) that is used in gesture determination and duration of phrase (used to decide which gesture macro to use, when to use it and how to set blink pattern)

4.3 Emotional Model

The emotion model in CBI has several unique features required to realize expressive, interactive psychosocial drama that set it apart from standard models of emotions used in agent systems. Central to these features is the fact that emotions in CBI are not simply there to make the characters more believable. Emotions and how individuals cope with emotional stress were an

integral aspect of the pedagogy. The pedagogical goal was for the mothers to learn how to choose and carry out the right coping strategy for a given situation and to maintain a realistic belief in their efficacy.

Consistent with the pedagogical role of emotion in CBI, the agents' interactions with each other and the learner are in fact grounded in the research that influenced the Bright IDEAS pedagogy: the cognitive appraisal theory of human emotion as posited by Richard Lazarus [25]. This theory organizes human behavior around appraisal and coping. Appraisal leads to emotion by assessing the person-environment relationship. This assessment is performed along several key dimensions. For example, did an event facilitate or inhibit the agent's goals; who deserves blame or credit? Most notable for the discussion here is the dimension of ego involvement: how an event impacts an individual's ego-identity. Ego-identity is the individual's collection of concerns for self- and social-esteem, social roles, moral values, self-ideals as well as concern for other people's well-being.

Coping is the process of dealing with emotion, either by acting externally on the world (problem-focused coping), or by acting internally to change beliefs or attention (emotion-focused coping). For example, a problem-focused way to attempt to deal with a loved one's illness is to take action that gets them medical attention. Alternatively, one might use an emotion-focused strategy such as avoiding thinking about it, focussing on the positive (e.g., one's love for an ill child) or denying the seriousness of the event. In Lazarus's theory, coping and appraisal interact and unfold over time, supporting the temporal character of emotion evident in human behavior.

As the previous tantrum example reveals, ego-identity and coping are key aspects of the emotion modeling in CBI. The focus on ego identity, in particular, distinguishes CBI from systems like Émile that model emotions that arise from tasks. In CBI, the knowledge modeled by the agent's ego identity comprises a key element of how it interacts with other characters and its response to events. For example, it is Carmen's concern for her son's well-being that induces sadness. And it is her ideal of being a good mother, and desire to be perceived as one (social esteem), that leads to anxiety about discussing Diana's tantrums with Gina.

Another key concern for CBI was to support the temporal character of emotion: an agent may "feel" distress for an event which motivates the shifting of blame, which leads to anger. The venting of that anger may in turn lead to guilt. In particular, capturing and expressing these dynamics in CBI lead to a design whereby agents emotionally evaluate and react to their own dialog. The expression of emotion also stems from two sources, appraisals [16] as well as the intention to communicate emotions [26, 39] that is derived from the current dialog act. Thus an agent can communicate emotions that they do not "feel". We will discuss in the Behavior Generation section how these two sources are mediated.

In CBI, ego-identity is modeled as a collection of role ideals (Carmen wants to be a good-mother), concerns (good-mothers want their children to

be happy and healthy) and responsibilities (good-mothers are responsible for their child's behavior). The system also models social relations (Gina is in essence a parental-surrogate for Carmen). Appraisal rules derive emotions from these various representations. Figure 8 describes some of the knowledge of ego-identity, roles and social relationships. Some of the appraisal rules used in the emotional processing are exemplified in Figure 9. Note that the appraisals are also performed on topic changes. Topics such as Diana's tantrums and Jimmy's illness have pre-existing emotional state information that is averaged into current emotion state when the topic is raised (as we will see MRE realizes such a capability in a more principled fashion). Emotions are represented as scalars on key types of emotion and coping factors. The appraisals result in changes in these values, which in turn impact dialog transitions, dialog rules and expressive behavior. Although there was an attempt to write these appraisal rules in a general fashion, the coverage is also partial, driven by the demands of the interactive story and characters and the pragmatic demand of getting CBI ready for clinical trials with real mothers.

Roles and Ideals

- (ego-ideal <person> <role> <type>)
  - Example: (ego-ideal Carmen mother good-mother)
- (concern <type> <relationship> <state>)
  - Example: (concern good-mother dependent positive–affect)
- (responsibility <type> <relationship> <state>)
  - Example: (responsibility good-mother dependent behavior)

Relationships

- (parental-surrogate <person> <person>)
  - Example: (parental-surrogate Carmen Gina)

Fig. 8. Example Ego-Identity Representations

4.4 CBI Dialog Model

The dialog model used by the CBI agents is designed to support considerable flexibility, dialog turn by dialog turn, while supporting interesting dramatic outcomes. Most interesting from an expressive behavior standpoint, all this flexibility in dialog is often driven by emotions, specifically the agents' emotional state, their coping strategies and their assessment of the other's agent's emotional state. To better appreciate this impact of emotions, we will briefly describe how dialog is generated.

The agent's dialog module selects high-level strategies to drive the discourse through the scene. These strategies are descriptions of possible realizations of the major components of the discourse and are designed to support

- If event violates a concern, it is negative.
- If asked by parental–surrogate about negative event which agent feels responsible for then increase anxiety.
- If talking about negative event then increase sadness.
- If talking about negative event which agent feels responsibility for then increase guilt.
- If new topic is raised and agent has pre-existing emotional attitude towards it then average in emotions with current emotional state.

Fig. 9. Example Appraisal Rules

considerable flexibility in the agent's turn-by-turn dialog. For example, the main act in CBI is Gina's goal of getting Carmen to apply the IDEAS steps to one of her problems. Gina has an abstract strategy to do this: **reassure** Carmen, **suggest** they jointly apply Bright IDEAS, **ask** her to choose a problem and **guide** her through the task of solving that problem - specifically guide her through the subgoals of I-D-E-A-S applied to that problem. This particular strategy sets an overall direction for the scene. The agents also have alternative substrategies that can hierarchically expand a strategy. For example the I-D-E-A-S subgoals need to be expanded. One substrategy is to repeatedly prompt/help the other agent to enumerate possible solutions to a subgoal. For example, Gina might use this substrategy to help Carmen **d**evelop (the D in IDEAS) possible solutions to Diana's tantrums. Another is to ask an ordered sequence of questions on a topic. Gina, for example, might help Carmen **i**dentify (I) the current problem's features by answering the "5Ws": who is at the center of the current problem being discussed, what is the problem, where does the problem happen, when does it happen and why does it happen.

The high level strategy and substrategies are not fixed prescriptions for the dialog. Rather, the agent expands the hierarchy and works out steps in the strategy interactively with the other agent. In the case of CBI, the expansion is done via joint agreement of Gina and Carmen. Gina suggests a substrategy like the "5Ws" and Carmen decides whether to agree to that approach. Each step in a strategy may need to be further expanded by the agents, via selecting another substrategy to expand it. Alternatively, a step may be primitive in the sense that there is no strategy to expand it. Such primitive steps are not single dialog turns, however. Rather, the agent generates its dialog turn-by-turn by flexibly interpreting the high-level strategies using a state machine. This machine allows the agent to adapt to twists and turns in the dialog caused by the autonomy of the agents and the learner's interactions. In the case of non-primitive steps in a strategy, it manages the dialog interactions which will

hopefully lead to an agreement on how to expand the step. Similarly, in the case of primitive steps, like answering the "why" question of the "5-Ws", the state machine manages how the agent will interact with the other agent to satisfy the step.

The state machine includes two kinds of nodes: dialog acts that generate a dialog turn and nodes that step through the current (sub)strategy being interpreted (e.g., Next Step). Both of these node types manage the dialog state by expanding a strategy, maintaining what the current strategy and topic is, where the agent is in the strategy and dialog obligations. Transitions occur between nodes depending on the current strategy, the current state of the dialog as well as the agent's and listener agent's emotional state. The dialog acts are:

- Suggest (e.g., a joint subplan),
- Agree (to subplan),
- Ask/Prompt (e.g., for an answer),
- Re-Ask/Re-Prompt,
- Answer or re-answer,
- Reassure (e.g., to impact listener's emotional state),
- Agree/Sympathize (convey sympathy),
- Praise,
- Offer-Answer (without being asked),
- Clarify (elaborate),
- Resign (give-up) and
- Summarize

Most notable are the ones that are tightly coupled to emotional state and pedagogy: Reassure, Praise, Agree/Sympathize, Resign (Give-up) and Summarize.

This design allows for both deliberative dialog and reactive dialog that variabilizes the agent interactions at multiple levels. At the highest level, alternative strategies and substrategies can be selected. Further, the specific transitions and resulting acts realize those strategies dialog turn by dialog turn in flexible ways, because a single step of the strategy can be realized by different paths through the agent's dialog state machine. Finally, there are typically multiple realization rules to address a specific act. For example, there may be multiple ways for Carmen to answer a specific question. Some of these may be qualitatively different in the sense that they lead to different recorded dialog lines and different dialog annotations. Such differences lead to a different resulting state of the system. Others may have the same resulting annotations, but actually use different lines or even the same line spoken with different affect. For example, Carmen has multiple ways to say many of her lines, using different affect (frustrated, depressed, optimistic, etc.), that are selected based on her emotional state.

Emotion and its expression play a key role in the dialog in other ways. For example, Gina's transitions between dialog acts are based on Carmen's

emotional state. She reassures or sympathizes when Carmen is distraught but prompts Carmen to address the current step in the current dialog strategy when Carmen is less distraught. If Carmen's emotion model leads her to respond inappropriately, Gina has to decide how to repair this failure. In psychological terms, Gina is often choosing whether to direct Carmen towards emotion-directed versus problem-directed coping by giving either emotional or instrumental support. Coping is key to the agent's selection of dialog and its response to it. Carmen may choose an evasive coping strategy and select dialog consistent with that strategy using the coping annotations. For example, the Carmen agent's emotion model appraises the discussion of Diana's tantrums as a source of distress because of her concern for Diana and because failure to control Diana may reflect on her ability as a mother. Her response to this stress may be to blame Diana and trivialize her tantrums by saying she is just being babyish. The Gina agent will not accept this answer, again because of the coping strategy annotation, and will ask a follow-on question. But Carmen may also reject her own answer first. Specifically if she is not too anxious or angry, the guilt caused by the answer may cause her to re-answer it prior to Gina's further prompting.

4.5 Behavior Generation

As noted, nonverbal behaviors are generated by the behavior generation module. The design of this module was heavily influenced by the psychological research of Freedman [15]. Freedman described behavior of clinical patients in terms of modes mediated by emotional state. In our computational model, we have delineated three modes: body-focus, transitional and communicative, roughly based on his work. These modes are arranged in a finite state machine, which we call a physical focus model. Body-focus mode is marked by a self-focused attention, away from the conversation and the problem-solving behavior. Emotionally, it is associated with considerable depression or guilt. Physically, it is associated with the tendencies of gaze aversion, paused or inhibited verbal activity and hand-to-body stimulation that is either soothing (e.g., rhythmic stroking of forearm) or self-punitive (e.g., squeezing or scratching of forearm). The agent doesn't exhibit communicative gestures such as deictic or beat gestures when in this mode. Transitional indicates a less withdrawn attention, less anxiety, a burgeoning willingness to take part in the conversation, milder conflicts with the problem solving and a closer relation to the listener. Physically, it can be marked by hand-to-hand fidgeting. There are more communicative gestures in this mode but they are still muted. Finally, communicative indicates a full willingness, or intent, to engage in the dialog and problem solving. Physically, it is marked by the agent's full range of communicative gestures and use of gaze in turn taking.

Behavior generation selects behavior based on physical focus mode. At any point in time, the agent will be in a specific mode based on emotional state that predisposes it to use nonverbal behavior in a particular fashion.

Each behavior available to an agent is categorized according to which subset of these modes it is consistent with. Any specific nonverbal behavior, such as a particular nod of the head, may exist in more than one mode, and conversely a type of behavior, such as head nods in general, may be realized differently in different modes. Transitions between modes are based on emotional state.

By grouping behaviors into modes, the physical focus mode attempts to mediate competing communicative and non-communicative demands on an agent's physical resources, in a fashion consistent with emotional state. Gestures, gaze and head movements are in particular driven by the physical focus mode. As we will see, facial expressions have a more temporal relation to the focus mode, driven by a desire to balance the expression of underlying emotional state with the communicative intent to express emotion as a social signal. This grouping model is designed to be general across agents. However, realism also requires that behaviors within each mode incorporate individual differences, as in human behavior. For example, Carmen's and Gina's repertoire of gestures incorporate individual differences.

Based on the current focus mode and emotional state, behavior rules, triggered in concert with the dialog and appraisal processes noted above, build a behavior program that expresses how those processes are unfolding. Behaviors include a combination of posture, head movement, facial expressions, blinking and dialog, arranged in an XML structure.

The structure of the behavior program consists of animation directives for the pieces of the agent's body, composed by parallel <P> and sequential <S> markers as well as pause animation directives. This allows recursive structures capable of simultaneous, sequential and delayed behaviors of arbitrary complexity. The XML structure in CBI is similar to other XML based animation languages, including most recently the work of Cassell et al. [8] and Pelachaud et al. [37]. Even though CBI's parallel, sequential and pause language is simple and quite aged now, it is somewhat unique in its ability to support timing of one behavior in absolute time or relative to any other behavior. Often XML languages only support timing of behaviors tied to the schedule of the speech.

Figure 10 depicts the high level XML structure of the resulting animation program for one phrase of an agent's dialog turn. Each box in the diagram would in turn be realized by nested XML animation directives. Note there are starting and ending expressions for the entire dialog turn as well as expressions that are displayed as the dialog turn unfolds, phrase by phrase. This allows the agent's behavior to reflect unfolding emotional signals driven by the multiple appraisals and sources of emotions as noted earlier. In particular, the starting facial expressions are driven by appraisal of the previous speaker's dialog. Expressions during the phrase are driven either by appraisal or the intent to communicate emotion derived by the dialog model. In communicative mode, it is the latter while in body and transitional mode it is the former. Note, the system originally used a weighted average of these two sources of emotion to select the expression but, in actual practice, there was insufficient expressive facial behavior in the animation resources to support

such averaged distinctions, a point we return to in section 4.6. Gestures are created in parallel with the phrase. Again the gestures used are determined by the focus mode, as well as the dialog annotations, including the referential structure noted earlier.

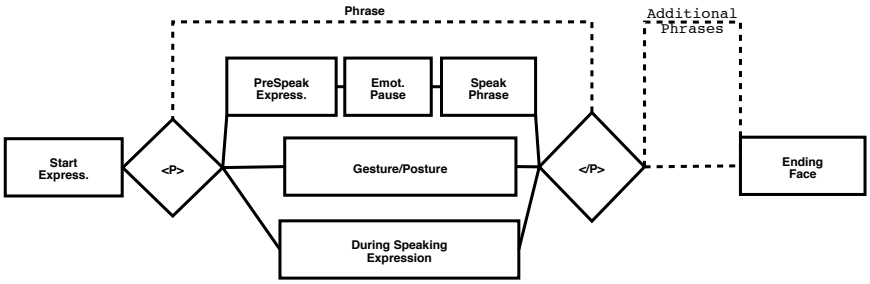

Fig. 10. High level description of XML structure for one phrase of an agent's dialog turn.

Each individual behavior, such as the starting emotional expression, may have a recursive parallel and sequential structure. Figure 11 depicts just one such expression, a guilt expression, that would be one small component of a full behavior program for a dialog turn. Note the expression involves two sequential threads of parallel motion involving eye, brow, mouth and head movements. The initial segment includes movement of the eyebrows and eyes. Once these animations are done, there is then a pause followed by final movements of head, mouth, eyes and brows. This use of structured and animated movements of the head even for just one component of the full animation program grew out of a small empirical study which suggested that expressive facial behavior was best revealed by changes in expression, including head movements, as opposed to static expressions.

Finally, note each animation directive has an "identifier". This is simply a pointer to a frame or sequence of frames in a vector animation file for that body part. The bodies used for Carmen and Gina were composed of separate parts, including legs, torso (posture), arms, head position, eyes, brows and mouth. Each of these assets were hand animated by artists and then stored in multi-frame animation files. This allowed the animation directives to identify single frames, say for a fixed arm position, or a sequence of frames for a particular movement of the arm.

4.6 Remarks

In Carmen's Bright IDEAS, the design of dialog, emotions and expressive behavior systems was driven by a need to convey deep inner conflicts and

```
<S>
  <P>
   (Carmen, brows, <identifier>)
   (Carmen, eyes, <identifier>)
  </P>
   (pause, <ticks>)
  <P>
   (Carmen,head, <identifier>)
   (Carmen,mouth, <identifier>)
   (Carmen,eyes, <identifier>)
   (Carmen,brows, <identifier>
  </P>
</S>
```

Fig. 11. Facial and head behavior for a guilt expression.

how those conflicts play out expressively over time. This lead the design of the agents along certain paths: how to model ego-identity, coping and the dynamics of expression. Perhaps the best test of its success in doing so were the trials with real mothers of pediatric cancer patients.

Carmen was evaluated in an exploratory arm of a larger clinical trial of the Bright IDEAS technique. The evaluation was very promising for Carmen and the use of interactive pedagogical drama in health interventions. Details on the results can be found in [30]. Overall, mothers were very enthusiastic. They found the story and its presentation in animated form to be very believable as well as a very effective and concrete way to learn Bright IDEAS. At the same time, one mother noted slowness in the graphics and other mothers wanted more story content that addressed other issues (such as marital concerns).

For the discussion in this chapter, this reveals a fundamental concern. One of the key issues that was faced in design of CBI was not having the sufficient resources to create the necessary dialog and animation assets to reveal all the expressive capabilities of the underlying models. Going forward with such applications will require us to leverage existing character animation frameworks as opposed to building our own as we did in Carmen. Nevertheless, systems like CBI could fill a void in making effective health intervention training available to the larger public. The training task for CBI was a difficult one, fraught with potentially many pitfalls. The fact that it was received so well by the mothers was remarkable and bodes well for applying IPD to other training and learning tasks.

5 Mission Rehearsal Exercise

The Mission Rehearsal Exercise (MRE) system [46] brings together ideas from each of the preceding systems to create a broader and more flexible array of expressive behaviors. The goal of the MRE is to teach leadership skills in

high-stakes social situations. The system places a human learner in command of a team of virtual humans interacting in an emotionally charged virtual environment. For example, in our initial scenario, the learner's team has, by accident, critically injured a young boy and the learner must juggle how to treat the boy without jeopardizing his mission or the safety of his team. To accomplish this, the learner must engage in face-to-face dialogue with his team members, take stock of the situation, give orders, and monitor their execution. Complicating this are characters that may express intense emotion and offer potentially biased or misleading information.

To model such dramatic and interactive scenarios, the MRE combines a number of elements of the preceding systems. We build on Steve's ability to flexibly interact with a human user, but augment it with the richer social and emotional behaviors of CBI and Jack and Steve. This has pushed us towards a tight integration of approaches, in some cases significantly altering their character, while at other times forcing us to defer key capabilities of the preceding systems for future research.

Although this chapter focuses on expressive behavior and the cognitive processes that support this, the MRE combines a variety of capabilities in service of realistic and natural collaboration with virtual humans including:

- a realistic model of human auditory and visual perception [44]
- a domain-specific finite-state speech recognizer that recognizes thousands of distinct utterances in noisy environments
- a finite-state semantic parser that produces (partial) semantic representations of the information in the text strings returned from speech recognition
- a dialogue model that explicitly represents aspects of the social context [49, 32] while supporting multi-party conversations and face-to-face communication in 3D virtual worlds [47]
- a dialogue manager that recognizes dialogue acts from utterances, updates the dialogue model, and selects new content for the virtual human to say
- a natural language generator that can produce nuanced English expressions, depending on the virtual human's personality and emotional state as well as the selected content [14]
- an expressive speech synthesizer capable of speaking in different voice modes depending on factors such as proximity (speaking vs. shouting) and illocutionary force (command vs. normal speech) [22]

The MRE seeks to advance the state of the art in each of these areas, but also to explore how best to integrate them into a single agent architecture [44], incorporating a flexible blackboard architecture to facilitate experiments with the connections between the individual components. We refer the reader to the above citations for details on these other components and here focus on our innovations in expressive behavior.

Figure 12 illustrates a scene from the MRE scenario. The learner plays the role of a lieutenant in the U.S. Army involved in a peacekeeping operation

in Bosnia. In route to assisting another unit, one of the lieutenant's vehicles becomes involved in a traffic accident, critically injuring a young boy. The boy's mother is understandably distraught and a local crowd begins to gather. The learner must resolve the situation by interacting through spoken dialogue with virtual humans in the scene.

Fig. 12. A scene from the MRE scenario.

5.1 Cognition and Emotions

To support such emotionally dramatic situations, virtual humans not only must produce a range of realistic expressive behaviors, but require the cognitive machinery to recognize which behaviors are appropriate in the course of an unscripted interaction with a human user. When selecting an expressive behavior, the virtual human's mental processes must take into account not only the task and dialogue context, as in the Steve system, but also how features of the social context will influence emotional appraisal and coping strategies.

We adopted Steve as a starting point for our integrated model of the cognition that underlies expressive behavior as, unlike Jack and Steve and CBI, the system supported flexible face-to-face interactions with a human user. However, Steve's task model had to be extended in a number of ways to represent the socio-emotional context.

Integrating the Émile plan-based appraisal model into Steve was relatively straightforward as both systems used similar task representations. Steve already possessed a model of task responsibility that supported appraisals of causal attribution, though Steve's task model had to be extended to represent probabilities and utilities and the processes that update these values. Steve also did not explicitly represent threats between task steps, necessary for appraisals of fear or anger, so we incorporated standard threat detection and resolution schemes.

One key difference between Steve and the Jack and Steve system was that the former focused on collaborative task execution and did not represent multiple competing ways to accomplish a task. (Steve would represent multiple alternative plans, but the current state of the world would always uniquely identify a "best" plan that all agents would be in agreement with.) We generalized the Steve task representation to encode multiple competing courses of action (recipes) for accomplishing a task. This allows agents to negotiate over tasks and express emotions and coping behaviors that may indicate varying preferences over alternatives [48].

Integrating CBI's coping model motivated further changes and resulted in a tight integration between appraisal, coping, and task reasoning that closely follows the cognitive appraisal theories of Richard Lazarus [25], and, in the end, elevated emotion processing to a central organizing construct for the virtual human's behavior. Recall this theory posits that human behavior is organized around appraisal and coping. Appraisal generates emotion by assessing the person-environment relationship and coping is the process of dealing with emotion, either by acting externally on the world (problem-focused coping), or by acting internally to change beliefs or attention (emotion-focused coping). Coping and appraisal interact and unfold over time, supporting the temporal character of emotion highlighted by CBI: an agent may "feel" distress for an event (appraisal), which motivates the shifting of blame (coping), which leads to anger (re-appraisal).

Through integrating CBI's coping model, coping strategies were recast explicitly into procedures that updated Steve's task representations and reasoning processes. For example, a strategy like problem-focused coping might motivate the task reasoner to refine its plans or motivate the agent to propose a particular course of action to the learner. Emotion-focused strategies like denial or shifting blame operate on the task representations, influencing the assignment of responsibility or altering the probability or utility of task consequences [29]. As many coping responses relate to past actions or decisions, we found it necessary to extend Steve's task representation to explicitly encode a causal history of past events and actions. Thus, appraisal and coping operate over a unified representation of past, present, and future task-related information.

Some of the key innovations of CBI and Jack and Steve have not, as of yet, been integrated into the MRE system, including planning stances and CBI's emphasis on ego identity. Although the integrated system does allow more flexibility in the plan generation process than the original Steve system, we have not adopted the full generative planning approach underlying Jack and Steve that planning stances require. In terms of cognitive appraisal, both CBI and Lazarus' theories emphasize the importance of ego identity. However, given Émile's heavy emphasis on task-related appraisals and domain-independent appraisal mechanisms, we have not found a suitable general way to represent the fact that certain threats are more central to an agent's makeup than others.

5.2 Physical Behavior

Internally, the virtual humans are continually perceiving the events surrounding them, understanding utterances, updating their beliefs, formulating and revising plans, generating emotional appraisals, and choosing actions. Virtual humans in the MRE attempt to manifest the rich dynamics of this cognitive and emotional inner state through each character's external behavior using the same verbal and nonverbal cues that people use to understand one another. The key challenge is the range of behaviors that must be seamlessly integrated: each character's body movements must reflect its awareness of events in the virtual world, its physical actions, the myriad of nonverbal signals that accompany speech during social interactions (e.g., gaze shifts, head movements, and gestures), and its emotional reactions. Expressive physical behavior in the MRE agents integrates the task-related nonverbal behaviors of the Steve system and the coping behaviors of CBI, leveraging the close integration of task-related and social information maintained by the virtual human's mental state.

Our use of gaze illustrates this tight integration. Since gaze indicates a character's focus of attention, it is a key element in any model of outward behavior, and must be closely synchronized to the character's inner thoughts. Prior work on gaze in virtual humans has considered either task-related gaze [9] or social gaze [5] but has not produced an integrated model of the two. Our gaze model is driven by our cognitive model, which interleaves task-related behaviors, social behaviors, and attention capture. Task-related behaviors (e.g., checking the status of a goal or monitoring for an expected effect or action) trigger a corresponding gaze shift, as does attention capture (e.g., hearing a new sound in the environment). Gaze during social interactions is driven by the dialogue state and the state of the virtual human's own processing, including gaze at an interlocutor who is speaking, gaze aversion during utterance planning (to claim or hold the turn), gaze at an addressee when speaking, and gaze when expecting someone to speak. This tight integration of gaze behaviors to our underlying cognitive model ensures that the outward attention of the virtual humans is synchronized with their inner thoughts.

Body movements are also critical for conveying emotional changes, including facial expressions, gestures, posture, gaze and head movements. In humans, these behaviors are signals and as such they can be used intentionally by an individual to inform or deceive but can also unintentionally reveal information about the individual's internal emotional state. Thus a person's behavior may express anger because they feel it or because they want others to think they feel it or for both reasons. With the exception of CBI, prior work on emotional expression in virtual humans focused on either the intentional use of emotional expression or revealing the agent's "true" internal emotional state [33]. Our work attempts to integrate these aspects by tying expressive behavior to coping behavior, generalizing the mechanism used in CBI. Emotional changes in the virtual human unfold as a consequence of Soar

operators updating the task representation. These operators provide a focus for emotional processes, invoking coping strategies to address the resulting emotions which in turn leads to expressive behaviors. This focus on operators both centers emotional expression on the agent's current internal cognitive processing but also allows coping to alter the relation of the expression to those internal cognitive processes. Thus, when making amends, our virtual humans might freely express their true appraisal-based feelings of guilt and concern, for example through facial expressions, gestures, posture, gaze and head movements. However, when shifting responsibility, they might suppress an initial expression of guilt and rather express anger at the character they are blaming, to reflect a more calculated attempt to persuade others.

Finally, a wide range of body movements are typically closely linked to speech, movements that emphasize, augment and even supplant components of the spoken linguistic information. Consistent with this close relation, this nonverbal behavior, which can include hand-arm gestures, head movements and postural shifts, is typically synchronized in time with the speech. Realizing this synchronization faces the challenge that we do not have an incremental model of speech production. Such a model would allow us to tie nonverbal behaviors to speech production operations much like the gaze and coping behaviors are tied to cognitive operations. Rather, our approach is to build on the gesture scheduling approach developed for CBI, which plans the utterance out and annotates it with nonverbal behavior. The annotated utterance is then passed to a text-to-speech generation system that schedules both the verbal and nonverbal behavior, using the BEAT system [8]. This approach is similar to the work of Cassell et al. [5]. Our work differs in the structure passed to the gesture annotation process, in order to capture the myriad ways that the nonverbal behavior can relate to the spoken dialog and the internal state of the virtual human. Specifically, while both systems pass the syntactic, semantic and pragmatic structure of the utterance, we additionally pass the emotional appraisal and coping information associated with the components of the utterance. The gesture annotation process uses this information to annotate the utterance with gestures, head movements, eyebrow lifts and eye flashes.

Some key aspects of Carmen's Bright IDEAS have not been incorporated into the current MRE system. CBI made effective use of the dramatic impact of pauses in speech, which can convey emotional turmoil or deliberation. CBI agents also had a far richer repertoire of expressive behaviors, particularly variability in motions associated with the eyes, eyelids, and brows. Such expressivity is not currently possible with the speech and animation systems used in the MRE system. While the MRE uses more realistic graphical models than the preceding three systems, they were developed by a third-party vendor, so we had less creative control over the animation than the other systems, which were developed in-house. Further the stylized 2D animation used in CBI supported a greater range of recognizable expressions. The greater complexity of MRE's natural language modules also limits the range of ex-

pressive behavior. In contrast to CBI, which used voice actors, MRE utilizes a fully automated speech generation pipeline, which provides the capability of dynamically generating a wide range of utterances, but allows far less nuanced speech, both in terms of emotional dynamics and creative use of pauses.

6 Conclusion

This chapter has shown the evolution of our ideas on expressive behaviors and their integration in our current virtual humans. Steve's ability to interleave task-related behaviors and face-to-face dialogue in dynamic virtual worlds serves as the foundation for the virtual humans in our MRE system. The Jack and Steve system contributed a model of task-oriented emotional appraisal (Émile) and a model of socially situated planning. The CBI system contributed a complementary model of emotional appraisal focusing on social relationships and ego identity, as well as a model of coping and of the effect of emotions and coping on physical behavior. Our MRE virtual humans integrate many of the ideas from these three prior systems, while significantly extending our prior work in some areas, such as our model of coping. The animation and speech capabilities in these four systems have offered different tradeoffs in generality and expressivity, illustrating the fact that any implemented model of expressive behavior must be closely integrated with the animation and speech capabilities available to it; otherwise, it may not be possible to accurately express the distinctions in that model. In our current work, we are continuing to push the frontiers of both our model of expressive behavior and its connection to the latest technologies in animation and speech production.

Acknowledgement. The Office of Naval Research funded the original research on Steve under grant N00014-95-C-0179 and AASERT grant N00014-97-1-0598. Lewis Johnson contributed to the original design of Steve, Randy Stiles and his colleagues developed the graphics software, Allen Munro and his colleagues developed the simulator, and Ben Moore helped link Steve to speech recognition.

The work on Carmen's Bright IDEAS was supported in part by the National Cancer Institute under grant R25CA65520. Our colleagues Lewis Johnson and Kate LaBore contributed significantly to the project.

The Department of the Army funds the MRE project under contract DAAD 19-99-D-0046. We thank our many colleagues who are contributing to the MRE project: Shri Narayanan leads a team working on speech recognition. Randy Hill, Mike van Lent, Changhee Han, and Youngjun Kim are working on models of agent perception. Ed Hovy, Deepak Ravichandran, and Michael Fleischman are working on natural language understanding and generation. Lewis Johnson, Kate LaBore, Shri Narayanan, and Richard Whitney are working on speech synthesis. Larry Tuch wrote the MRE story line with creative input from Richard Lindheim and technical input on Army procedures from Elke Hutto and General Pat O'Neal. Sean Dunn, Sheryl Kwak, Ben Moore, and Marcus Thiébaux created the simulation infrastructure for MRE. Marcus also developed the character animation system for Steve and Jack and

Steve. Any opinions, findings, and conclusions expressed in this article are those of the authors and do not necessarily reflect the views of the Department of the Army.

References

1. Jose A. Ambros-Ingerson and Sam Steel. Integrating planning, execution and monitoring. In *Proceedings of the Seventh National Conference on Artificial Intelligence (AAAI-88)*, pages 83 88, San Mateo, CA, 1988. Morgan Kaufmann.
2. L. Berkowitz. *Causes and Consequences of Feelings.* Cambridge University Press, 2000.
3. Rama Bindiganavale, William Schuler, Jan M. Allbeck, Norman I. Badler, Aravind K. Joshi, and Martha Palmer. Dynamically altering agent behaviors using natural language instructions. In *Proceedings of the Fourth International Conference on Autonomous Agents*, pages 293–300, New York, 2000. ACM Press.
4. Justine Cassell, Tim Bickmore, Lee Campbell, Hannes Vilhjálmsson, and Hao Yan. Conversation as a system framework: Designing embodied conversational agents. In J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, editors, *Embodied Conversational Agents.* MIT Press, Cambridge, MA, 2000.
5. Justine Cassell, Catherine Pelachaud, Norman Badler, Mark Steedman, Brett Achorn, Tripp Becket, Brett Douville, Scott Prevost, and Matthew Stone. Animated conversation: Rule-based generation of facial expression, gesture and spoken intonation for multiple conversational agents. In *Proceedings of ACM SIGGRAPH '94*, pages 413–420, Reading, MA, 1994. Addison-Wesley.
6. Justine Cassell, Joseph Sullivan, Scott Prevost, and Elizabeth Churchill, editors. *Embodied Conversational Agents.* MIT Press, Cambridge, MA, 2000.
7. Justine Cassell and Kristinn R. Thórisson. The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. *Applied Artificial Intelligence*, 13:519–538, 1999.
8. Justine Cassell, Hannes Vilhjálmsson, and Timothy Bickmore. Beat: the behavior expression animation toolkit. In *Proceedings of ACM SIGGRAPH*, pages 477–486, New York, 2001. ACM Press.
9. Sonu Chopra-Khullar and Norman I. Badler. Where to look? Automating attending behaviors of virtual human characters. *Autonomous Agents and Multi-Agent Systems*, 4(1-2):9–23, 2001.
10. Philip R. Cohen and Hector J. Levesque. Teamwork. *Nous*, 25(4):487–512, 1991.
11. Marie desJardins and Michael J. Wolverton. Coordinating a distributed planning system. *AI Magazine*, 20(4):45–53, Winter 1999.
12. Starkey Duncan, Jr. Some signals and rules for taking speaking turns in conversations. In Shirley Weitz, editor, *Nonverbal Communication*, pages 298 311. Oxford University Press, 1974.
13. Clark Elliott. *The Affective Reasoner: A Process Model of Emotions in a Multi-agent System.* PhD thesis, Northwestern University, 1992.
14. Michael Fleischman and Eduard Hovy. Emotional variation in speech-based natural language generation. In *Proceedings of the International Natural Language Generation Conference*, New York, 2002. Arden House.
15. Norbert Freedman. The analysis of movement behavior during the clinical interview. In A.W. Siegman and B. Pope, editors, *Studies in Dyadic Communication*, pages 177–210. Pergamon Press, New York, 1972.

16. Jonathan Gratch. Émile: Marshalling passions in training and education. In *Proceedings of the Fourth International Conference on Autonomous Agents*, pages 325–332, New York, 2000. ACM Press.
17. Jonathan Gratch. Socially situated planning. In Lola Cañamero Kerstin Dautenhahn, Alan H. Bond and Bruce Edmonds, editors, *Socially Intelligent Agents: Creating Relationships with Computers and Robots*, pages 181 188. Kluwer Academic Publishers, 2002.
18. Jonathan Gratch and Stacy Marsella. Tears and fears: Modeling emotions and emotional behaviors in synthetic agents. In *Proceedings of the Fifth International Conference on Autonomous Agents*, pages 278–285, New York, 2001. ACM Press.
19. B. J. Grosz and S. Kraus. Collaborative plans for complex group action. *Artificial Intelligence*, 86(2):269 357, 1996.
20. Barbara J. Grosz and Candace L. Sidner. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3):175 204, 1986.
21. W. Lewis Johnson, Jeff W. Rickel, and James C. Lester. Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial Intelligence in Education*, 11:47–78, 2000.
22. W.L. Johnson, S. Narayanan, R. Whitney, R. Das, M. Bulut, and C. LaBore. Limited domain synthesis of expressive military speech for animated characters. In *Proceedings of the IEEE Workshop on Speech Synthesis*, Santa Monica, CA, 2002.
23. Subbarao Kambhampati, Craig A. Knoblock, and Qiang Yang. Planning as refinement search: A unified framework for evaluating design tradeoffs in partial-order planning. *Artificial Intelligence*, 76:167–238, 1995.
24. John E. Laird, Allen Newell, and Paul S. Rosenbloom. Soar: An architecture for general intelligence. *Artificial Intelligence*, 33(1):1–64, 1987.
25. R.S. Lazarus. *Emotion and Adaptation*. Oxford Press, 1991.
26. James C. Lester, Stuart G. Towns, Charles B. Callaway, Jennifer L. Voerman, and Patrick J. FitzGerald. Deictic and emotive communication in animation pedagogical agents. In J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, editors, *Embodied Conversational Agents*. MIT Press, Cambridge, MA, 2000.
27. James C. Lester, Jennifer L. Voerman, Stuart G. Towns, and Charles B. Callaway. Deictic believability: Coordinating gesture, locomotion, and speech in lifelike pedagogical agents. *Applied Artificial Intelligence*, 13:383–414, 1999.
28. James C. Lester, Luke S. Zettlemoyer, Joel Gregoire, and William H. Bares. Explanatory lifelike avatars: Performing user-designed tasks in 3d learning environments. In *Proceedings of the Third International Conference on Autonomous Agents*, New York, 1999. ACM Press.
29. Stacy Marsella and Jonathan Gratch. Modeling coping behavior in virtual humans: Don't worry, be happy. In *Proceedings of the Second International Joint Conference on Autonomous Agents and Multi-Agent Systems*. ACM Press, 2003.
30. Stacy Marsella, W. Lewis Johnson, and Catherine LaBore. Interactive pedagogical drama for health interventions. In *Proceedings of the 11th International Conference on Artificial Intelligence in Education*. IOS Press, 2003.
31. Stacy C. Marsella, W. Lewis Johnson, and Catherine LaBore. Interactive pedagogical drama. In *Proceedings of the Fourth International Conference on Autonomous Agents*, pages 301–308, New York, 2000. ACM Press.

32. Colin Matheson, Massimo Poesio, and David Traum. Modelling grounding and discourse obligations using update rules. In *Proceedings of the First Conference of the North American Chapter of the Association for Computational Linguistics*, 2000.
33. W. Scott Neal Reilly. *Believable Social and Emotional Agents.* PhD thesis, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, 1996. Technical Report CMU-CS-96-138.
34. Allen Newell. *Unified Theories of Cognition.* Harvard University Press, Cambridge, MA, 1990.
35. A. Ortony, G.L. Clore, and A. Collins. *The Cognitive Structure of Emotions.* Cambridge University Press, 1988.
36. Catherine Pelachaud, Norman I. Badler, and Mark Steedman. Generating facial expressions for speech. *Cognitive Science*, 20(1), 1996.
37. Catherine Pelachaud, Valeria Carofiglio, Berardina De Carolis, Fiorella de Rosis, and Isabella Poggi. Embodied contextual agent in information delivering application. In *Proceedings of the First International Joint Conference on Autonomous Agents and Multi-Agent Systems*, pages 758 765, New York, 2002. ACM Press.
38. J. Penley and J. Tomaka. Associations among the big five, emotional responses, and coping with acute stress. *Personality and Individual Differences*, 32:1215–1228, 2002.
39. Isabella Poggi and Catherine Pelachaud. Emotional meaning and expression in performative faces. In *International Workshop on Affect in Interactions: Towards a New Generation of Interfaces*, Siena, Italy, 1999.
40. Jeff Rickel and W. Lewis Johnson. Animated agents for procedural training in virtual reality: Perception, cognition, and motor control. *Applied Artificial Intelligence*, 13:343 382, 1999.
41. Jeff Rickel and W. Lewis Johnson. Virtual humans for team training in virtual reality. In *Proceedings of the Ninth International Conference on Artificial Intelligence in Education*, pages 578–585. IOS Press, 1999.
42. Jeff Rickel and W. Lewis Johnson. Task-oriented collaboration with embodied agents in virtual worlds. In J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, editors, *Embodied Conversational Agents.* MIT Press, Cambridge, MA, 2000.
43. Jeff Rickel and W. Lewis Johnson. Extending virtual humans to support team training in virtual reality. In G. Lakemayer and B. Nebel, editors, *Exploring Artificial Intelligence in the New Millenium*, pages 217–238. Morgan Kaufmann, San Francisco, 2002.
44. Jeff Rickel, Stacy Marsella, Jonathan Gratch, Randall Hill, David Traum, and William Swartout. Toward a new generation of virtual humans for interactive experiences. *IEEE Intelligent Systems*, 17(4):32–38, 2002.
45. K.R. Scherer. Appraisal considered as a process of multilevel sequential checking. In K.R. Scherer, A. Schorr, and T. Johnstone, editors, *Appraisal Processes in Emotion: Theory, Methods, and Research*, pages 92–120. Oxford University Press, 2001.
46. W. Swartout, R. Hill, J. Gratch, W.L. Johnson, C. Kyriakakis, C. LaBore, R. Lindheim, S. Marsella, D. Miraglia, B. Moore, J. Morie, J. Rickel, M. Thiébaux, L. Tuch, R. Whitney, and J. Douglas. Toward the holodeck: Integrating graphics, sound, character and story. In *Proceedings of the Fifth International Conference on Autonomous Agents*, pages 409 416, New York, 2001. ACM Press.

47. David Traum and Jeff Rickel. Embodied agents for multi-party dialogue in immersive virtual worlds. In *Proceedings of the First International Joint Conference on Autonomous Agents and Multi-Agent Systems*, pages 766–773, New York, 2002. ACM Press.
48. David Traum, Jeff Rickel, Jonathan Gratch, and Stacy Marsella. Negotiation over tasks in hybrid human-agent teams for simulation-based training. In *Proceedings of the Second International Joint Conference on Autonomous Agents and Multi-Agent Systems*. ACM Press, 2003.
49. David R. Traum. *A Computational Theory of Grounding in Natural Language Conversation*. PhD thesis, Department of Computer Science, University of Rochester, Rochester, NY, 1994.
50. David R. Traum and Elizabeth A. Hinkelman. Conversation acts in task-oriented dialogue. *Computational Intelligence*, 8(3):575 599, 1992.
51. Daniel S. Weld. An introduction to least commitment planning. *AI Magazine*, 15(4):27 61, 1994.

Index