Embodied Conversational Agents, J. Cassell, S. Prevost, and J. Sullivant (Eds.). Boston: MIT Press.

Deictic and Emotive Communication in Animated Pedagogical Agents

James C. Lester Stuart G. Towns Charles B. Callaway
Jennifer L. Voerman Patrick J. FitzGerald

The IntelliMedia Initiative
North Carolina State University

1 Introduction

Lifelike animated agents offer great promise for knowledge-based learning environments. Because of the immediate and deep affinity that children seem to develop for these agents, the potential pedagogical benefits they provide may perhaps even be exceeded by their motivational benefits. By creating the illusion of life, animated agents may significantly increase the time that children seek to spend with educational software. Recent advances in affordable graphics hardware are beginning to make the widespread distribution of real-time animation technology a reality, so children across the socioeconomic spectrum will reap its benefits. Endowing animated agents with believable, lifelike qualities has been the subject of a growing body of research (André and Rist, chap. XX; Badler, chap. XX; Bates 1994; Blumberg and Galyean 1995; Cassell et al. 1994, Kurlander and Ling 1995). Researchers have begun to examine the incorporation of gesture and facial expression in embodied conversational agents (Cassell, chap. XX; Poggi and Pelachaud, chap. XX), and the social aspects of human-computer interaction and users' anthropomorphization of software (Isbister and Nass 1998) has been the subject of increasing interest in recent years.

Animated pedagogical agents (Rickel and Johnson, chap. XX; Lester, Stone, and Stelling 1999; Paiva and Machado 1998) constitute an important category of animated agents whose intended use is educational applications. A recent large-scale, formal empirical study suggests that these agents can be pedagogically effective (Lester et al. 1997b), and it was determined that students perceived the agent as being very helpful, credible, and entertaining (Lester et al. 1997a). Although these results are preliminary and precise measures of agents' pedagogical contributions will begin to appear as the technologies mature and longitudinal studies are undertaken, we believe the *potential* for animated pedagogical agents is significant. The work described here is part of a long-term research program to bring about fundamental improvements in learning environments by broadening the bandwidth of "face-to-face" tutorial communication (Johnson, Rickel, and Lester, in press).

Designing engaging animated pedagogical agents that communicate effectively involves a broad and complex matrix of psycholinguistic and engineering phenomena, many of which are discussed in this book. Two of these issues, *deictic believability* and *emotive believability*, are particularly important for animated pedagogical agents that are (1) situated in (virtual representations of) physical worlds that they immersively co-inhabit with students, and (2) designed to engage students affectively.

In the same manner that humans refer to objects in their environment through combinations of speech, locomotion, and gesture, animated agents should be able to move through their environment, point to objects, and refer to them appropriately as they provide problem-solving advice. Deictic believability in animated agents requires the design of an agent behavior planner that considers the physical properties of the world inhabited by the agent. The agent must exploit its knowledge of the positions of objects in the world, its relative location with respect to these objects, as well as its prior explanations to create deictic gestures, motions, and utterances that are both natural and unambiguous.

In addition to deictic believability, animated pedagogical agents should also exhibit emotive believability. Engaging lifelike pedagogical agents that are visually expressive can clearly communicate problem-solving advice and simultaneously have a strong motivating effect on students. Drawing on a rich repertoire of emotive behaviors to exhibit contextually appropriate facial expressions and expressive gestures, they can exploit the visual channel to advise, encourage, and empathize with students. However, enabling lifelike pedagogical agents to communicate the affective content of problem-solving advice poses serious challenges. Agents' full-body emotive behaviors must support expressive

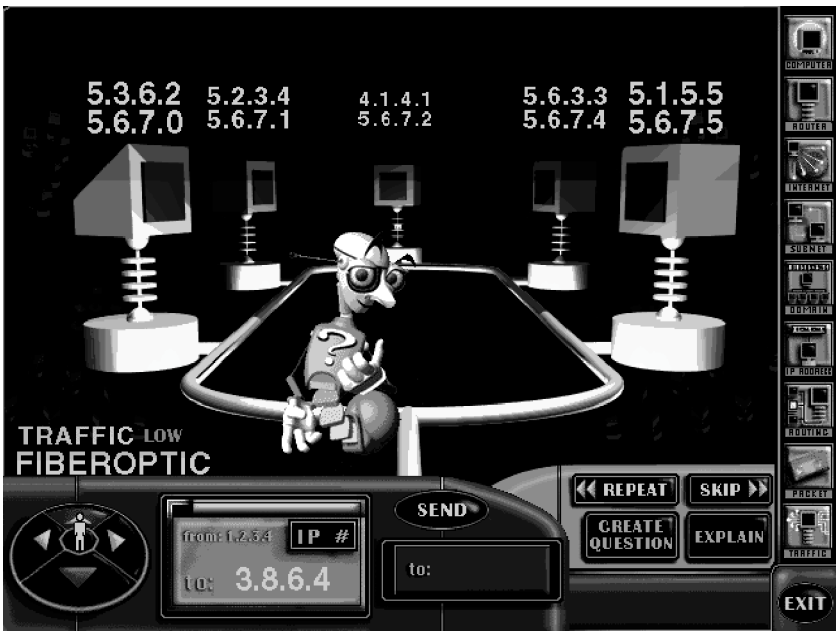

Figure 1. Cosmo and the Internet Advisor world.

movements and visually complement the problem-solving advice they deliver. Moreover, these behaviors must be planned and coordinated in real time in response to students' progress. In short, to create the illusion of emotional life typified by well-crafted animated characters, animated pedagogical agents must be able to communicate through both visual and aural channels.

To address these issues, we have developed the *spatial deixis* framework for achieving deictic believability and the *emotive-kinesthetic behavior sequencing* framework for dynamically sequencing lifelike pedagogical agents' full-body emotive expression. In the spatial deixis framework, a deictic behavior planner exploits a world model and the evolving explanation plan as it selects and coordinates locomotive, gestural, and speech behaviors. In the emotive-kinesthetic behavior sequencing framework, a behavior space is populated with emotive behaviors with full-body movements, including facial expressions featuring eye, eyebrow, and mouth movements as well as gestures with arms and hands; these behaviors are produced in real time by sequencing pedagogical speech acts and their associated emotional intent and kinesthetic expression. These frameworks have been used to implement Cosmo (Figure 1), a lifelike pedagogical agent with real-time deictic planning and full-body emotive expression. Cosmo inhabits the Internet Advisor, a learning environment for the domain of Internet packet routing. An impish, antenna-bearing creature who

hovers about in the virtual world of routers and networks, he provides advice to students as they decide how to ship packets through the network to specified destinations.

2 Deictic Behavior Sequencing

In the course of communicating with one another, interlocutors employ deictic techniques to create context-specific references. Hearers interpret linguistic events in concrete contexts. To understand a speaker's utterance, hearers must consider the physical and temporal contexts in which the utterance is spoken, as well as the identities of the speaker and hearer. Referred to as the *deictic center* of an utterance, the trio of location, time, and identities also plays an important role in generating linguistic events (Fillmore 1975). The first of these, location, is critical for achieving *spatial deixis*, a much-studied phenomenon in linguistics that is used to create references in the physical world (Jarvella and Klein 1982). Speakers use spatial deixis to narrow hearers' attention to particular entities. In one popular framework for analyzing spatial deixis, the *figure-ground* model (Roberts 1993), the world is categorized into *ground*, which is the common physical environment shared by the speaker and hearer, and the *referent*, the aspect of the ground to which the speaker wishes to refer. Through carefully constructed referring expressions and well-chosen gestures, the speaker assists the hearer in focusing on the particular referent of interest.

The ability to handle spatial deixis effectively is especially critical for animated pedagogical agents that inhabit virtual worlds. To provide problem-solving advice to students who are interacting with objects in the world, the agent must be able to refer to objects in the world to explain their function clearly and to assist students in performing their tasks. Deictic mechanisms for animated pedagogical agents should satisfy three criteria:

1. *Lack of Ambiguity:* In a learning environment, an animated agent's clarity of expression is of the utmost importance. To effectively communicate advice and explanations to students, the agent must be able to create deictic references that are unambiguous. Avoiding ambiguity is critical in virtual environments, where an ambiguous deictic reference can cause mistakes in problem solving and foster misconceptions. Ambiguity is particularly challenging in virtual environments housing a multitude of objects, especially when many of the objects are visually similar.

2. *Immersivity:* An agent's explanatory behaviors should be situated (Suchman 1987), that is, all of its actions—not merely its advisory actions but also its communication of conceptual knowledge—should take place in concrete problem-solving contexts. Frequently, these are (virtual) physical contexts. For example, in the course of delivering problem-solving advice, an agent frequently needs to refer to a particular object; it should be able to combine speech, gesture, and locomotion immersively (i.e., within a 2D or 3D environment) to do so, for example, by walking across a scene to a cluster of objects and pointing to one of them as its makes a verbal reference to the object.

3. *Pedagogical Soundness:* Deictic mechanisms for agents that inhabit learning environments must support their central pedagogical intent. Rather than operating in a communicative vacuum, spatial deixis must support the ongoing advisory discourse and be situated appropriately in the problem-solving context.

The lack-of-ambiguity requirement implies that deictic planning mechanisms must make use of an expressive representation of the world. While unambiguous deictic references can be created with object highlighting or by employing a relatively stationary agent with a long pointer (André and Rist 1996), the immersivity requirement implies that lifelike agents should artfully combine speech, gesture, and locomotion. Finally, the pedagogical soundness requirement implies that all deictic utterances, speech, and movements must be integrated with explanation plans that are generated in response to student questions and problem-solving impasses.

Following the lead of Bates (1994), we refer to the *believability* of lifelike agents as the extent to which users interacting with them come to believe that they are observing a sentient being with its own beliefs, desires, intentions, and personality. It has been shown that believable pedagogical agents in interactive learning environments can produce *the persona effect*, in which the very presence of a lifelike character in a learning environment can have a strong positive effect on students' perception of their learning experience (Lester et al. 1997a). A study involving one hundred middle school students revealed that when they interact with a lifelike agent that is expressive—namely, an agent that exhibits both animated and verbal advisory behaviors-students perceive it to be encouraging and useful.

A critical but largely unexplored aspect of agents' believability for learning environments is deictic believability. We say that lifelike agents making deictic

references in a manner that achieves a lack of ambiguity, that does so in an immersive setting, and that operates in a pedagogically sound manner exhibit *deictic believability*.

2.1 Related Work in Deictic Generation

The natural language generation and intelligent multimedia communities have addressed several aspects of spatial deixis. Natural language researchers have studied reference generation, for example, Dale's classic work on referring expressions (Dale 1992), scene description generation (Novak 1987), and spatial layout description generation (Sibun 1992). Work on intelligent multimedia systems (André et al. 1993; Feiner and McKeown 1990; Maybury 1991; Mittal et al. 1995; Roth, Mattis, and Mesnard 1991) has produced techniques for dynamically incorporating highlights, underlines, and blinking (Neal and Shapiro 1991). However, none of these consider the orchestration of an agent's communicative behaviors in an environment.

Work on lifelike agents has yielded more sophisticated techniques for referring to on-screen entities. The Edward system (Claassen 1992) employs a stationary persona that "grows" a pointer to a particular object in the interface, and the PPP agent (André and Rist 1996) is able to indicate dynamically various on-screen objects with a long pointer. While these techniques are effective for many tasks and domains, they do not provide a general solution for achieving deictic believability that deals explicitly with ambiguity by both selecting appropriate referring expressions and by producing lifelike gestures and locomotion.

Begun at the University of Pennsylvania's Jack project and continued at MIT, the work of Cassell and colleagues on conversational agents is perhaps the most advanced to date on agents that combine gesture, speech, and facial expression (Cassell et al. 1994). In addition to deictics, they also exhibit iconic, metaphoric, and beat gestures. However, this work neither provides a solution to the intricacies of detecting ambiguity in complex physical environments (and then addressing it with integrated speech, gesture, and locomotion) nor focuses on pedagogical interactions.

Despite the promise of lifelike pedagogical agents, with the exception of work on the Design-A-Plant project (Lester, Stone, and Stelling 1999) and the Soar Training Expert for Virtual Environments (Steve) (Rickel and Johnson, chap. XX), in which agents provide instruction about procedural tasks in a virtual

reality environment, lifelike agents for pedagogy have received little attention. Neither the Steve nor the Design-A-Plant projects address deictic believability.

2.2 A Deictic Believability Test Bed

Features of environments, agents, and tasks that force spatial deixis issues to the forefront are threefold: (1) A world populated by a multitude of objects, many of which are similar, will require agents to plan speech, gesture, and locomotion carefully to avoid ambiguity. (2) We can select a domain and problem-solving task for students that requires agents to provide advice and explanations that frequently refer to different objects in the world. (3) Problem-solving tasks that require students to make decisions based on factors physically present in the environment will induce clarity requirements on agents' communicative capabilities. In contrast to a more abstract domain such as algebra, we can select a domain that can be represented graphically with objects in perhaps idiosyncratic and complex spatial layouts, thereby requiring the agent to produce clear problem-solving advice that integrates spatial deixis with explanations of concepts and problem-solving strategies.

To investigate deictic believability in lifelike pedagogical agents, we have developed a test bed in the form of an interactive learning environment. Because it has each of the features outlined above, the Internet Advisor provides a "laboratory" in which to study the coordination of deictic speech, gesture, and locomotion. Designed to foster exploration of computational mechanisms for animation behavior sequencing of lifelike characters and real-time human-agent problem-solving interaction, the Internet Advisor consists of a virtual world populated by many routers and networks.

Students interact with Cosmo as they learn about network routing mechanisms by navigating through a series of subnets. Given a packet to escort through the Internet, they direct it through networks of connected routers. At each subnet, they may send their packet to a specified router or view adjacent subnets. By making decisions about factors such as address resolution and traffic congestion, they learn the fundamentals of network topology and routing mechanisms. Helpful, encouraging, and with a bit of attitude, Cosmo explains how computers are connected, how routing is performed, what types of networks have different physical characteristics, how Internet address schemes work, and how network outages and traffic considerations come into play. Students' journeys are complete when they have navigated the network successfully and delivered their packet to its proper destination. The learning environment serves as an excellent test bed for

exercising spatial deixis because each subnet has a variety of routers attached to it and the agent must refer unambiguously to them as it advises students about their problem-solving activities.

2.3 Coordinating Deictic Gesture, Locomotion, and Speech

The primary role of lifelike pedagogical agents is to serve as an engaging vehicle for communication. Hence, in the course of observing a student attempt different solutions in a learning environment, a lifelike pedagogical agent should clearly explain concepts and convey problem-solving strategies. It is in this context that spatial deixis arises. The spatial deixis framework guides the operation of the *deictic planner*, a key component of the agent behavior planning architecture. The interaction manager provides an interface between the learning environment and the agent that inhabits it. By monitoring a student's problem-solving activities in the learning environment, the interaction manager invokes the agent behavior planner in two situations: (1) when a student pauses for an extended period of time, which may signal a problem-solving impasse, and (2) when a student commits an error, which indicates a possible misconception.

The agent behavior planner consists of an explanation planner and a deictic planner. The explanation planner serves an analogous function to that of the discourse planner of natural language generation systems (Hovy 1993; Lester and Porter 1997; Moore 1995; Suthers 1991). Natural language generation systems typically consist of a discourse planner that determines the content and structure of multisentential texts and a realization system that plans the surface structure of the resulting prose. Analogously, given a communicative goal, the explanation planner of the agent behavior planner determines the content and structure of an agent's explanations and then passes these specifications to the deictic planner, which realizes these specifications in speech, gesture, and locomotion. The explanation planner invokes the deictic planner by specifying a communicative act, a topic, and a referent.

To accomplish its task, the deictic behavior planner examines the representational structures in a world model, a curriculum information network, a user model, the current problem state (which includes both the student's most recently proposed solution and the learning environment's analysis of that solution), and two focus histories, one for gesture and one for speech. (Algorithmic details of the deictic behavior planner can be found in Lester et al. 1999.) It then constructs a sequence of physical behaviors and verbal explanations that will collectively constitute the advice that the agent will deliver. For example, given a communicative goal, the

explanation planner for Cosmo typically produces an explanation plan that calls for the agent to speak from six to ten utterances and perform several locomotive and gestural behaviors. These are then passed to the presentation manager that manipulates the agent persona in the learning environment. Problem-solving actions performed by the student are therefore punctuated by customized explanations provided by the agent in a manner reminiscent of classic task-oriented dialogues.

Deictic planning comes into play when the behavior planner determines that an explanation must refer to an object in the environment. For each utterance that makes a reference to an environmental object, the explanation planner invokes the deictic system and supplies it with the intended referent. The deictic system operates in the following phases to plan the agent's gestures, locomotion, and speech:

1. *Ambiguity Appraisal:* The deictic system first assesses the situation by determining whether a reference may be ambiguous. By examining the evolving *explanation plan*, which contains a record of the objects the agent has referred to during utterances spoken so far in the current explanation sequence, the deictic planner evaluates the initial potential for ambiguity. This assessment will contribute to gesture, locomotion, and speech planning decisions.

2. *Gesture and Locomotion Planning:* The deictic system uses the specification of the relative positions of the objects in the scene of the world model, as well as the previously made ambiguity assessment, to plan the agent's deictic gestures and movement. By considering the proximity of objects in the world, the deictic system determines whether the agent should point to the referent and, if so, whether it should move to it.

3. *Utterance Planning and Coordination:* To determine what the agent should say to refer to the referent, the deictic system considers focus information, the ambiguity assessment, and the world model. Utterance planning pays particular attention to the relative locations of the referent and the agent, taking into account its planned locomotion from the previous phase. The result of utterance planning is a referring expression consisting of the appropriate proximal/nonproximal demonstratives and pronouns. Finally, the behavior planner coordinates the agent's spoken, gestural, and locomotive behaviors,

orchestrates their exhibition by the agent in the learning environment, and returns control to the student.

The computational methods underlying ambiguity appraisal, gesture and locomotion planning, and deictic referring expression planning are described below.

2.3.1 Ambiguity Appraisal

The first phase of deictic planning consists of evaluating the potential for ambiguity. For each utterance in the evolving explanation plan that makes a reference to an object in the environment, the explanation planner invokes the deictic system. Deictic decisions depend critically on an accurate assessment of the discourse context in which the reference will be communicated. To plan the agent's gestures, movements, and utterances correctly, the deictic system determines whether the situation has the potential for ambiguity within the current explanation. This initial phase of ambiguity assessment considers only discourse issues; spatial considerations are handled in the two phases that follow. Because focus indicates the prominence of the referent at the current juncture in the explanation, the deictic system uses focus as the primary predictor of ambiguity: potentially ambiguous situations can be combated by combinations of gesture and locomotion.

A referent *R* has the potential for ambiguity if it is currently not in focus or if it is in focus but is one of multiple objects in focus. To determine if the referent is in focus, the deictic system examines the evolving explanation plan and inspects it for previous deictic references to *R*. Suppose the explanation planner is currently planning utterance U_i. It examines utterances U_{i-1} and U_{i-2} for preceding deictic references to *R*. There are three cases to consider:

1. *Novel Reference:* If the explanation planner locates no deictic reference to *R* in U_{i-1} or $U_{i-2,}$ then *R* is ambiguous and is therefore deserving of greater deictic emphasis. For example, if a student interacting with the Internet Advisor chooses to send a packet to a particular router that does not lie along the optimal path to the packet's destination, Cosmo interrupts the student and makes an initial reference to that router. He should therefore introduce the referent into the discourse.

2. *Unique Focus:* If the explanation planner locates a reference to *R* in U_{i-1} and U_{i-2} but not to other entities, then *R* has already been introduced and the potential for ambiguity is less. For example, when Cosmo's explanation consists of multiple utterances about a particular router, a reference to that router will be in unique focus. Consequently, the need for special deictic treatment is reduced.

3. *Multiple Foci:* If the explanation planner locates a reference to *R* but also to other entities in U_{i-1} and U_{i-2}, then the situation is potentially ambiguous. For example, if Cosmo points to one router and subsequently points to another that the student has just selected, but he now needs to refer to the first router again for purposes of comparison, multiple referents are in focus and he must therefore take precautions against making an ambiguous reference.

The result of this determination is recorded for use in the following two phases of gesture and locomotion planning and referring expression planning.

2.3.2 Gesture and Locomotion Planning

When potential ambiguities arise, endowing the agent with the ability to point and move to objects to which it will be referring enables it to increase its clarity of reference. The deictic system plans two types of physical behaviors: gestures and locomotion. In each case, it first determines whether a behavior of that type is warranted. If so, it then computes the behavior.

To determine whether the agent should exhibit a pointing gesture to designate physically the referent within the environment, the behavior planner inspects the conclusion of the ambiguity computation in the previous phase. If the referent was deemed ambiguous or potentially ambiguous, the system will plan a pointing gesture for the agent.

In addition to pointing, the agent can also move from one location to another to clarify a deictic reference that might otherwise be ambiguous. If the referent has been determined to be unambiguous—namely, it is in a unique focus—the agent will remain stationary. (More precisely, the agent will not perform a locomotive behavior; in fact, for purposes of believability, the agent is always in subtle but constant motion, for example, Cosmo typically performs "antigravity bobbing" and blinking behaviors.) In contrast, if the referent is ambiguous—that is, if it is a novel reference—the deictic system instructs the agent to move toward the object

specified by the referent as the agent points at it. For example, if Cosmo is discussing a router that has not been previously mentioned in the last two utterances, he will move to that router as he points to it. If the referent is potentially ambiguous, that is, it is a reference to one of the concurrently active foci, then the deictic planner must decide if locomotion is needed. If no locomotion is needed, the agent will point at *R* without moving toward it. In contrast, if any of the following three conditions hold, the agent will move toward *R* as it points:

1. *Multiple Proximal Foci:* If the object specified by *R* is near another object that is also in focus, the agent will move to the object specified by *R*. For example, if two nearby routers are being compared, Cosmo will move to the router to which he is referring to ensure that his reference is clear.

2. *Multiple Proximal Similarity:* Associated with each object is an ontological category. If the object specified by *R* is near other objects of the same category, the agent will move to the object specified by *R*. For example, if Cosmo were referring to a computer and there were several computers nearby, he would move to the intended computer.

3. *Diminutiveness*: If the object specified by *R* is unusually small, the agent will move to the object specified by *R*. Small objects are labeled as such in the world model. For example, many interface control buttons are relatively small compared to objects in the environment. If Cosmo needs to make a clear reference to one of them, he will move toward that button.

After a sequence of high-level gestures and locomotive behaviors are computed, they must be interpreted within the learning environment. For example, the current implementation of Cosmo provides for six basic pointing gestures: left-up, left-across, left-down, right-up, right-across, and right-down. To enable the agent to point correctly to the object specified by the referent, the behavior planner first consults the world model. It obtains the location of the agent (L_A) and the referent (L_R) in the environment. It then determines the relative orientation of the vector from (L_A) to (L_R). For example, Cosmo might be hovering in the lower-left corner of the environment and need to point to a router in the upper-right corner. In this case, he will point up and to his left toward the router using his left-up gesture.

The behavior planner must then determine whether or not the agent really needs to move based on his current location. If it determines that locomotion is called

for, the interaction manager must first determine if the agent is already near the object, which would obviate the need to move toward it. Nearness of two objects is computed by measuring the distance between them and ascertaining whether it is less than a *proximity bound*. If the distance between the agent and the intended object is less than the proximity bound, then there is no need for the agent to move because it can already point clearly to the object, and so it will remain in its current position.

If locomotion is appropriate, the behavior planner computes a direct motion path from the agent's current location to the object specified by *R*. To do so, it first determines the *deictic target*, which is the precise location in the world at which the agent will point. To avoid ambiguity, the agent will move its finger (or, more generally, its deictic pointer) toward the center of referent. It then computes the direction of the vector defining the agent's direction of travel from L_A and to the deictic target. To do so, it first determines the position of its finger if it were extended in the direction computed in step 2 with the agent in L_A. It then determines the ideal location of the agent's body position if its outstretched finger were to touch the deictic target in the final position. Finally, it traverses the resulting motion path connecting L_A and its final body position and location.

2.3.3 Deictic Referring Expression Planning and Coordination

To communicate effectively the intended reference, the deictic system must combine gesture, locomotion, and speech. Having completed gesture and locomotion planning, the deictic planner turns to speech. To determine an appropriate referring expression for the agent to speak as it performs the deictic gestures and locomotion, the deictic system first examines the results of the ambiguity appraisal. If it was determined that *R* is in unique focus, there is no potential for ambiguity because *R* has already been introduced and no other entities are competing for the student's attention. It is therefore introduced with a simple referring expression using techniques similar to those outlined in Dale (1992). For example, “the router” will be pronominalized to “it.”

In contrast, if *R* is ambiguous or potentially ambiguous—namely, *R* is a novel reference or is one of multiple foci—the deictic planner makes three assessments: (1) it determines the demonstrative category called for by the current situation; (2) it examines the ontological type of *R* and the other active foci; and (3) it considers the number of *R*. It first categorizes the situation into one of two deictic families:

1. *Proximal Demonstratives:* If the deictic planner has determined that the agent must move to *R* or that it would have moved to *R* if it were not already near *R*, then employ a proximal demonstrative such as "this" or "these."

2. *Nonproximal Demonstratives*: If the deictic planner has determined that *R* was not nearby but that the agent did not need to move to *R*, then employ a nonproximal demonstrative such as "that" or "those."

After it has determined which of the demonstrative categories to use, the deictic planner narrows its selection further by considering the ontological type of *R* and the previous two utterances in the evolving explanation plan. If *R* belongs to the same ontological type as the other entities that are in focus, then the deictic planner selects the phrase "This one . . ." For example, suppose the system has determined that a proximal demonstrative should be used and that the preceding utterance referred to one router, for example, "This router has more traffic." To refer to a second router in the current utterance, rather than saying, "This router has less traffic," it will say, "This one has less traffic." Finally, it uses the number of *R* to make the final lexical choice. If *R* is singular, it uses "this" for proximal demonstratives and "that" for nonproximals. If *R* is plural, it uses "these" and "those." The resulting referring expression is then passed onto the behavior planner for the final phase.

To integrate the agent's physical behaviors and speech, the behavior planner then coordinates the selected utterances, gestures, and locomotion. Three types of coordination must be achieved. First, each utterance may be accompanied by a deictic gesture, and it is critical that the agent's referring expressions be tightly coupled to its corresponding pointing movements. Second, pointing and locomotion should be carefully coordinated so that they occur in a natural manner, where "natural" suggests that the agent should perform its pointing gesture *en route* to the referent and arrive at the referent at precisely the same time that it reaches the apex of the pointing gesture. Third, when the agent exhibits a sequence of speech, gestural, and locomotive behaviors to communicate an explanation, the behavior planner must ensure that each cluster of utterances, gestures, and possible agent movements are completed before the next is initiated. The behavior planner enacts the coordination by specifying that the utterance be initiated when the agent reaches the apex of its pointing gesture. In contrast, if the speech were initiated at the same time as the gesture and locomotion, the utterance would seem to complete prematurely, thereby producing both ambiguity and the appearance of incongruous behavior.

Finally, to underscore the deictic gestures, the behavior planner introduces gaze into the final behavior. As demonstrated by Cassell's incorporation of a gaze generator in her conversational agents (Cassell et al. 1994), gaze offers an important communication medium for acknowledgments and turn taking. In addition, gaze can play an important role in deixis. For example, when Cosmo refers to a particular computer on a subnet by moving toward it and pointing at it as he speaks about it, he should also look at it. The behavior planner enacts gaze via specifications for the agent to "look" at the referent by moving its head in precisely the direction in which it is pointing. In the implementation, the behavior planner accomplishes this not through run-time inference of eye control but by exploiting agent head rendering in which the eyes were crafted by the animators to look in the direction in which the head is pointing, for example, if the head is turned toward the right, the eyes look toward the right. The USC/ISI animated agents group has been successfully experimenting with similar gaze techniques such as "leading with the eyes" (Johnson and Rickel 1997).

The behavior planner combines the speech, gesture, locomotion, and gaze specifications and directs the agent to perform them in the order dictated by the explanation plan. The agent's behaviors are then assembled and sequenced in the learning environment in real time to provide students with clear advice that couples full deictic expression with integrated lifelike locomotion, gesture, speech, and gaze. Currently, although the resulting behaviors are coordinated after they have been constructed, limited (and ad hoc) communication occurs between the modules that individually plan speech, gesture, and locomotion. An important direction for future work, which is currently being pursued in the Rea project (Cassell and Stone 1999), is a principled model of media allocation in which the individual modules can communicate bidirectionally with one another to carefully plan the content to be conveyed in each modality.

3 Emotive Behavior Sequencing

In the same manner that human-human communication is characterized by affective multimodal interaction utilizing both the visual and aural channels, agent-human communication can be achieved in a similar fashion. As master animators have discovered repeatedly over the past century, the quality, overall clarity, and dramatic impact of communication can be increased through the creation of emotive movement that underscores the affective content of the message to be communicated: by carefully orchestrating facial expression, full-body behaviors, arm movements, and hand gestures, animated pedagogical agents could visually augment verbal problem-solving advice, give encouragement,

convey empathy, and perhaps increase motivation. Although initial forays have begun on emotion generation in pedagogical environments (Abou-Jaoude and Frasson 1998) and reasoning about students' emotions (de Vicente and Pain 1998), emotive behavior sequencing in pedagogical agents remains unexplored.

Creating lifelike pedagogical agents that are endowed with facilities for exhibiting student-appropriate emotive behaviors potentially provides four important educational benefits (Elliott, Rickel, and Lester 1999). First, a pedagogical agent that appears to care about a student's progress may convey to the student that it and she are "in things together" and may encourage the student to care more about her own progress. Second, an emotive pedagogical agent that is in some way sensitive to the student's progress may intervene when she becomes frustrated and before she begins to lose interest. Third, an emotive pedagogical agent may convey enthusiasm for the subject matter at hand and may foster similar levels of enthusiasm in the student. Finally, a pedagogical agent with a rich and interesting personality may simply make learning more fun. A student who enjoys interacting with a pedagogical agent may have a more positive perception of the overall learning experience and may consequently opt to spend more time in the learning environment.

3.1 The Emotive-Kinesthetic Behavior Framework

To enable a lifelike pedagogical agent to play an active role in facilitating students' progress, its behavior sequencing engine must be driven by students' problem-solving activities. As students solve problems, an explanation system monitors their actions in the learning environment (Figure 2). When they reach an impasse, as indicated by extended periods of inactivity or suboptimal problem-solving actions, the explanation system is invoked to construct an explanation plan that will address potential misconceptions. By examining the problem state, a curriculum information network, and a user model, the explanation system determines the sequence of pedagogical speech acts that can repair the misconception and passes the types of the speech acts to the emotive-kinesthetic behavior sequencing engine. By assessing the speech act categories and then selecting full-body emotive behaviors that the agent can perform to communicate the affective impact appropriate for those speech act categories, the behavior sequencing engine identifies relevant behaviors and binds them to the verbal utterances determined by the explanation system. The behaviors and utterances are then performed by the agent in the environment and control is returned to the student who continues her problem-solving activities.

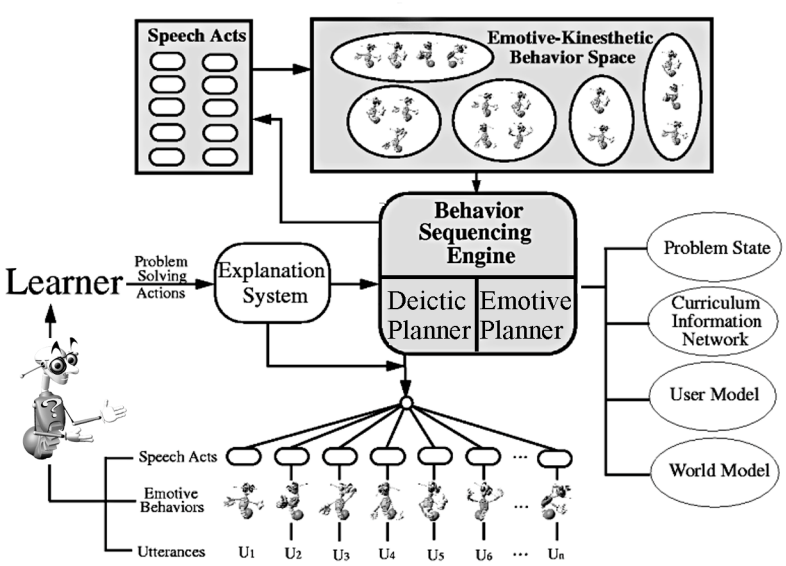

Figure 2. The lifelike pedagogical agent behavior planning architecture.

The techniques for designing emotive-kinesthetic behavior spaces, the representations for structuring them with pedagogical speech act categories, and the computational mechanisms that drive the emotive behavior sequencing engine are described below.

3.2 Emotive-Kinesthetic Behavior Space Design

To exhibit full-body emotive behaviors, a pedagogical agent's behavior sequencing engine must draw on a large repertoire of behaviors that span a broad emotional spectrum. For many domains, tasks, and target student populations, fully expressive agents are very desirable. To this end, the first phase in creating a lifelike pedagogical agent is to design an *emotive-kinesthetic behavior space* that is populated with physical behaviors that the agent can perform when called upon to do so. Because of the aesthetics involved, an agent's behaviors are perhaps best designed by a team that includes character animators. Creating a behavior space entails setting forth precise visual and audio specifications that describe in great detail the agent's actions and speech, rendering the actions, and creating the descriptive utterances. By exploiting the character behavior canon of the animated film (Jones 1989; Noake 1988) (which itself draws on movement in theater) and then adapting it to the specific demands posed by learning environments, we can

extract general emotive animation techniques that artists in this medium have developed over the past hundred years.

3.2.1 Stylized Emotive Behaviors

It is important to draw the critical distinction between two approaches to animated character realization, life-quality versus stylized (Culhane 1988). In the *life-quality* approach, character designers and animators follow a strict adherence to the laws of physics. Characters' musculature and kinesthetics are defined entirely by the physical principles that govern the structure and movement of human (and animal) bodies. For example, when a character becomes excited, it raises its eyebrows and its eyes widen. In the *stylized* approach, the laws of physics (and frequently the laws of human anatomy and physiology) are broken at every turn. When a character animated with the stylized approach becomes excited, for example, as in the animated films of Tex Avery (Lenburg 1993), it may express this emotion in an exaggerated fashion by rising from the ground, inducing significant changes to the musculature of the face, and bulging out its eyes. Not all stylized animation features such exaggerated emotive overstatement—for learning environments, a more restrained approach is called for—but its ability to communicate with dramatic visual cues can be put to good use in the real-time animation of pedagogical agents. For example, when a student solves a complex problem in the Internet Advisor environment, Cosmo smiles broadly and uses his entire body to applaud the student's success.

3.2.2 Expressive Range

To be socially engaging, animated characters must be able to express many different kinds of emotion. As different social situations arise, they must be able to convey emotions such as happiness, elation, sadness, fear, envy, shame, and gloating. In a similar fashion, because lifelike pedagogical agents should be able to communicate with a broad range of speech acts, they should be able to support these speech acts visually with an equally broad range of emotive behaviors. However, because their role is primarily to facilitate positive learning experiences, only a critical subset of the full range of emotive expression is useful for pedagogical agents. For example, they should be able to exhibit body language that expresses joy and excitement when students do well, inquisitiveness for uncertain situations (such as when rhetorical questions are posed), and disappointment when problem-solving progress is less than optimal. For example,

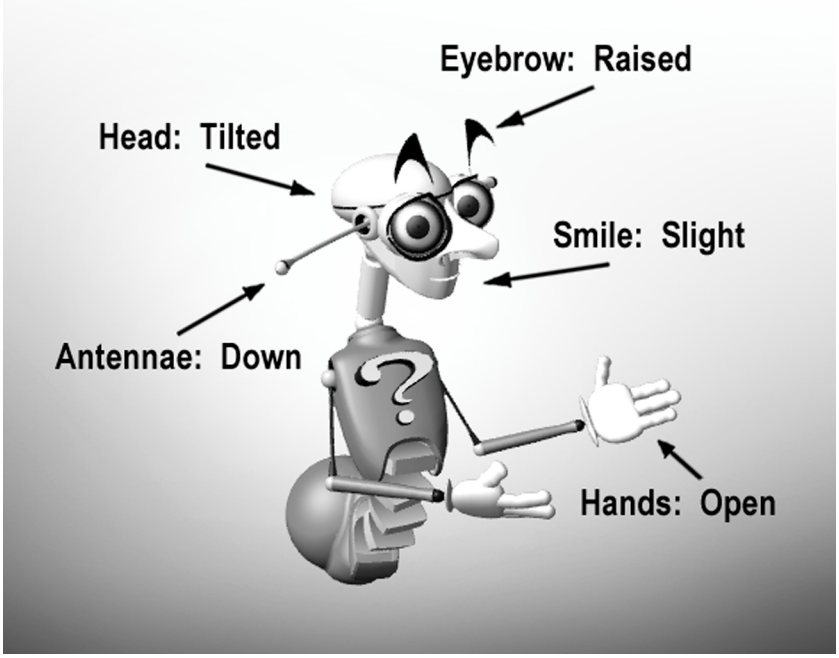

Figure 3. Sample Cosmo posture.

the Cosmo agent can scratch his head in wonderment when he poses a rhetorical question.

3.2.3 Anatomical Emotive Information Carriers

Years of experimentation in animation demonstrate that specific anatomical components communicate emotion more than others. By focusing on the more expressive components, we can create lifelike agents that convey emotive content more effectively. For example, longtime Disney animators stress the critical importance of the hands (Thomas and Johnston 1981). It is for this reason that great attention is paid to hand movement and that hands are often rendered much larger than would be anatomically correct. Although literally every body part can be used to convey emotion, the principle carriers of emotive information are the eyes, eyebrows, face, mouth, head tilt, posture, and gesturing with the arms and hands. For example, Figure 3 depicts a frame of Cosmo taken from a behavior in which he appears quizzically friendly. His eyebrows are raised, his head is slightly askew, his mouth forms a smile, and his hands are raised. Moreover, stylized characters can have additional appendages to further convey emotion. For example, in the frame of Cosmo shown in Figure 3, his antennae droop slightly.

3.3 Behavior Space Structuring with Pedagogical Speech Acts

An agent's behaviors will be dictated by design decisions in the previous phase, which to a significant extent determine its personality characteristics. Critically, however, its run-time emotive behaviors must be somehow modulated to a large degree by ongoing problem-solving events driven by the student's activities. Consequently, after the behavior space has been populated with expressive behaviors, it must then be structured to assist the sequencing engine in selecting and assembling behaviors that are appropriate for the agent's communicative goals. Although, in principle, behavior spaces could be structured along any number of dimensions such as degree of exaggeration of movement or by the type of anatomical components involved in movements, experience with the implemented agent suggests that the most effective means for imposing a structure is based on *speech acts*. While it could be indexed by a full theory of speech acts, our research to date leverages a highly specialized collection of speech acts that occur in pedagogical dialogue with great frequency.

Given the primacy of the speech act in this approach, the question then arises about the connection between rhetorical goals, on the one hand, and physical behaviors, on the other. Emotive categories inspired by foundational research on affective reasoning supply this linkage. Work on the Affective Reasoner (AR) (Elliott 1992) uses Ortony's computational model of emotion (Ortony, Clore, and Collins 1988) to design agents that can respond emotionally. In the AR framework, agents are given unique pseudopersonalities modeled as both an elaborate set of *appraisal frames* representing their individual goals (with respect to events that arise), *principles* (with respect to perceived intentional actions of agents), *preferences* (with respect to objects), *moods* (temporary changes to the appraisal mechanism), and as a set of about 440 differentially activated *channels* for the expression of emotions (Elliott 1992; Elliott and Ortony 1992). Situations that arise in the agents' world may map to twenty-six different emotion types (e.g., *pride*, as approving of one's own intentional action), twenty-two of which were originally theoretically specified by Ortony and his colleagues (Ortony *et al.*, 1988). Quality and intensity of emotion instances in each category are partially determined by some subset of roughly twenty-two different *emotion intensity variables* (Elliott and Siegle 1993). To communicate with users, Elliott's implementation of the AR framework uses line-drawn facial expressions, which are morphed in real time.

The emotive-kinesthetic behavior sequencing framework exploits the fundamental intuition behind the AR—namely, that the emotive states and communication are intimately interrelated. Rather than employing the full computational apparatus of

the AR, the emotive-kinesthetic framework uses highly simplified emotive annotations that connect pedagogical speech acts to relevant physical behaviors. Computationally, this is accomplished by employing a model of communication that places pedagogical speech acts in a *one-to-one* mapping to emotive states: each speech act type points to the behavior types that expresses it. To illustrate, in creating the Cosmo agent, the design team focused on the speech acts (and their associated emotions) that are prominent in problem-solving tutorial dialogues. The Cosmo agent deals with cause and effect, background, assistance, rhetorical links, and congratulatory acts as follows:

- *Congratulatory act*: When a student experiences success, a congratulatory speech act triggers an admiration emotive intent that will be expressed with behaviors such as applause, which, depending on the complexity of the problem, will be either restrained or exaggerated. The desired effect is to encourage the student.

- *Causal act:* When a student requires problem-solving advice, a causal speech act is performed in which the agent communicates an interrogative emotive intent that will be expressed with behaviors such as head scratching or shrugging. The desired effect is to underscore questioning.

- *Deleterious effect*: When a student experiences problem-solving difficulties or when the agent needs to pose a rhetorical question with unfortunate consequences, disappointment is triggered that will be expressed with facial characteristics and body language that indicate sadness. The desired effect is to build empathy.

- *Background* and *assistance*: In the course of delivering advice, background or assistance speech acts trigger inquisitive intent that will be expressed with "thoughtful" restrained manipulators such as finger drumming or hand waving. The desired effect is to emphasize active cognitive processing on the part of the agent.

The one-to-one mapping is used to enact a threefold adaptation of the AR framework. First, while the AR framework is intended to be generic, the emotive-kinesthetic behavior framework is designed specifically to support problem-solving advisory communication. Second, while the AR framework is enormously complex, the emotive-kinesthetic framework employs only the speech acts and only the emotive intentions that arise frequently in tutorial situations. Third, while work on computational models of social linguistics indicates that the combination of speech and gesture in human-human communication is enormously complex

(Cassell, chap. XX), the one-to-one mapping approach turns out in practice to be a reasonable starting point for real-time emotive behavior sequencing.

To create a fully operational lifelike agent, the behavior space includes auxiliary structuring to accommodate important emotive but non-speech-oriented behaviors such as dramatic entries into and exits from the learning environment. Moreover, sometimes the agent must connect two behaviors induced by multiple utterances that are generated by two speech acts. To achieve these rhetorical link behaviors, it employs subtle "micromovements" such as slight head nods or blinking.

3.4 Dynamic Emotive Behavior Sequencing

To dynamically orchestrate full-body emotive behaviors that achieve situated emotive communication, complement problem-solving advice, and exhibit real-time visual continuity, the emotive behavior sequencing engine selects and assembles behaviors in real time. By exploiting the pedagogical speech act structuring, the sequencing engine navigates coherent paths through the emotive behavior space to weave the small local behaviors into continuous global behaviors. Given a communicative goal G, such as explaining a particular misconception that arose during problem solving, a simple overlay user model, a curriculum information network, and the current problem state, it employs the following algorithm to select and assemble emotive behaviors in real time:

1. Determine the pedagogical speech acts $A_1 \ldots A_n$ used to achieve G. When the explanation system is invoked, employ a top-down goal decomposition planner to determine a set of relevant speech acts. For each speech act A_i, perform steps (2)–(5).

2. Identify a family of emotive behaviors F_i to exhibit when performing A_i. Using the emotive annotations in the behavior speech act structuring, index into the behavior space to determine a relevant family of emotive behaviors F_i.

3. Select an emotive behavior B_i that belongs to F_i. Either by using additional contextual knowledge, for example, the level of complexity of the current problem, or by choosing randomly when all elements of F_i are relevant, select an element of F_i.

4. Select a verbal utterance U_i from the library of utterances that is appropriate for performing A_i. Using an audio library of voice clips that is analogous to physical behaviors, extract a relevant voice clip.

5. Coordinate the exhibition of B_i with the speaking of U_i. Couple B_i with U_i on the evolving timeline schedule.

6. Establish visual continuity between B_1 ... B_n. Examine the final frame of each B_i, compare it with the initial frame of each B_{i+1}, and if they differ, introduce transition frames between them.

For each speech act A_i identified in step 1, the sequencing engine performs the following actions. During step 2, it identifies a family of emotive behaviors F_i that can be exhibited while the agent is performing A_i. It accomplishes this by employing pedagogical speech act indices that have been used to index the agent's physical behavior space. For example, a congratulatory speech act created during top-down planning will cause the sequencing engine to identify the admiration emotive behavior family.

Next, during step 3, it selects one of the physical behaviors in F_i. By design, all of the behaviors have the same emotive intent, so they are all legitimate candidates. However, because a key aspect of agent believability is exhibiting a variety of behaviors, the behavior space was constructed so as to enable the agent to perform a broad range of facial expression and gestures. Hence, the sequencing engine selects from a collection of behaviors, any of which will effectively communicate the relevant emotive content. For example, in the current implementation of the Cosmo agent, the behavior sequencing engine makes this decision pseudorandomly with elimination—that is, it randomly selects from among the behaviors in F_i that have not already been marked as having been performed. After all behaviors in a given F_i have been performed, they are unmarked, and the process repeats. Empirical evidence suggests that this pseudorandom element contributes significantly to believability.

During the final three steps, the behavior sequencing engine determines the narrative utterances to accompany the physical behaviors and assembles the specifications on an evolving timeline. In step 4, it selects the narrative utterances U_i, which are of three types: *connective* (e.g., "but" or "and"), *phrasal* (e.g., "this subnet is fast"), or *sentential* (i.e., a full sentence). Because each instantiated speech act specifies the verbal content to be communicated, narrative utterance selection is straightforward. In step 5, it lays out the physical behaviors and verbal utterances in tandem on a timeline. Because the emotive physical behaviors were

determined by the same computational mechanism that determined the utterances, the sequencing engine can couple their exhibition to achieve a coherent overall behavior.

Finally, in step 6, it ensures that the visual continuity is achieved by introducing appropriate transition frames. To do so, for each of the visual behaviors selected above, it inspects the first and final frames. If adjacent behaviors are not visually identical, it splices in visual transition behaviors and installs them, properly sequenced, into the timeline.

The sequencing engine passes all behaviors and utterances to the learning environment, which cues them up and orchestrates the agent's actions and speech in real time. The net effect of the sequencing engine's activities is the student's perception that an expressive lifelike character is carefully observing their problem-solving activities and behaving in a visually compelling manner. The resulting behaviors are then exhibited by the agent in the learning environment, and control is immediately returned to the student who continues her problem-solving activities.

4 An Implemented, Full-Body Emotive Pedagogical Agent

The spatial deixis framework and the emotive-kinesthetic framework have been implemented in Cosmo, the lifelike (stylized) pedagogical agent that inhabits the Internet Advisor learning environment. Cosmo and the Internet Advisor environment are implemented in C++, employ the Microsoft Game Software Developer's Kit (SDK), and run on a PC at 15 frames/second. Cosmo's deictic planner is implemented in the CLIPS production system language (NASA 1993).

Cosmo has a head with movable antennae and expressive blinking eyes, arms with bendable elbows, hands with a large number of independent joints, and a body with an accordionlike torso. His speech was supplied by a voice actor. Cosmo was modeled and rendered in 3-D on SGIs with Alias/Wavefront. The resulting bitmaps were subsequently postedited and transferred to PCs where users interacted with them in a $2^1/_2$-D environment. Cosmo can perform a variety of behaviors including locomotion, pointing, blinking, leaning, clapping, and raising and bending his antennae. His speech was created by a trained voice actor and an audio engineer. His verbal behaviors include 240 utterances ranging in duration from one to twenty seconds.

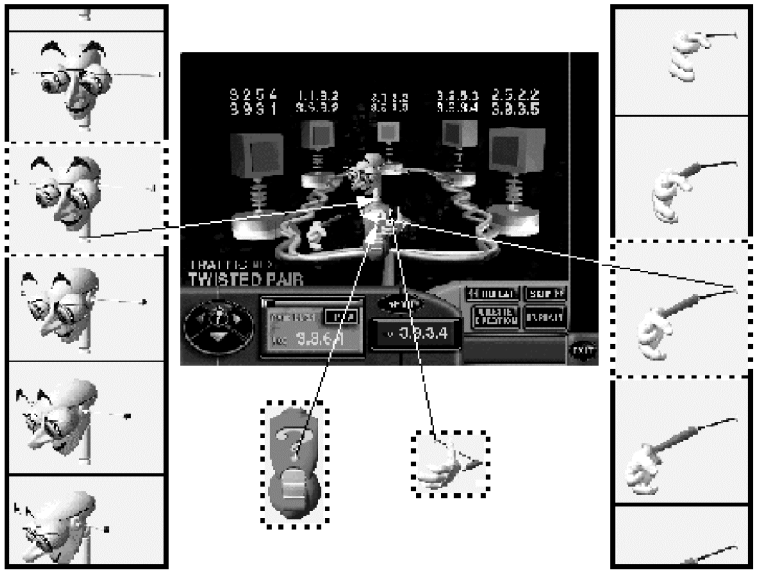

Figure 4. Real-time composition of deictic gestural and gaze components.

Cosmo's behaviors are assembled in real time as directed by the behavior planner. Each action is annotated with the number of frames and transition methods. Actions are of two types: *full-body* behaviors, in which the agent's entire body is depicted, and *compositional* behaviors that represent various body parts individually. To sequence nondeictic behaviors such as clapping and leaning, the behavior planner employs the full-body images. To sequence deictic behaviors, including both the gesture and gaze, the behavior planner combines compositional behaviors of torsos, left and right arms, and heads (Figure 4).

As the student attempts to route her packet to a given destination, she makes a series of routing decisions to direct the packet's hops through the network. At each router, she is given four different subnets, each with five possible computers with multiple unique addresses from which to choose. She is also provided information about the type of the subnet and the amount of traffic on the subnet. In the lower left-hand corner of the interface, she can click on different quadrants of a spinner to navigate among the four possible attached subnets. When she has found what she believes to be a reasonable computer toward which to send her packet, she clicks on the address of the computer. Cosmo then comments on the correctness and optimality of her decision. If it is either incorrect or suboptimal, he provides assistance on how to improve it. If her decision was deemed optimal,

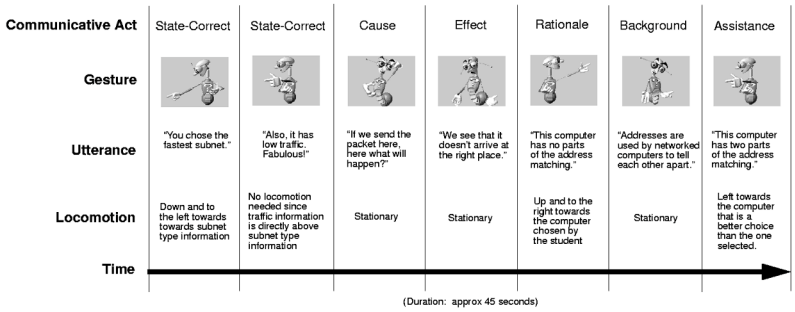

Figure 5. Sample behavior sequencing.

he congratulates her, and she clicks on the "Send" button to send her packet to the next subnet in the network.

To illustrate the behavior planner's behavior, suppose a student has just routed her packet to a fiber optic subnet with low traffic. She surveys the connected subnets and selects a router that she believes will advance it one step closer to the packet's intended destination. Although she has chosen a reasonable subnet, it is suboptimal because of nonmatching addresses, which will slow her packet's progress. She has made two mistakes on address resolution already, so the explanation is somewhat detailed. Working in conjunction, the deictic and emotive behavior planners select and sequence the following communicative acts and orchestrate the agent's gestural, locomotive, and speech behaviors (Figure 5):

- State-Correct(Subnet-Type): The explanation planner determines that the agent should interject advice and invokes the deictic planner. Since nothing is in focus because this is the first utterance to be planned for a new explanation, and Cosmo currently occupies a position on the screen far from information about the subnet—namely, the distance from his current location to the subnet information exceeds the proximity bound—he moves toward and points at the onscreen subnet information and says, "You chose the fastest subnet."

- State-Correct(Traffic): Cosmo then tells the student that the choice of a low traffic subnet was also a good one. The focus history indicates that while the type of subnet has already been the subject of a deictic reference, the traffic information has not. Cosmo therefore moves to the on-screen congestion information and points to it. However, the focus history indicates that he has mentioned the subnet in a recent utterance, so he pronominalizes the subnet as "it" and says, "Also, it has low traffic."

- Congratulatory(Generic): Responding to a congratulatory speech act, the sequencing engine selects an admiration emotive intent that is realized with an enthusiastic applauding behavior as Cosmo exclaims, “Fabulous!”

- Causal(Generic): The sequencing engine’s planner selects a *causal* speech act, which causes the interrogative emotive behavior family to be selected. These include actions such as head scratching and shrugging, for which the desired effects are to emphasize a questioning attitude. Hence, because Cosmo wants the student to rethink her choice, he scratches his head and poses the question, “But more importantly, if we sent the packet here, what will happen?”

- Deleterious-Effect(Address-Resolution): After the causal act, the sequencing engine’s planner now selects a deleterious-effect speech act, which causes it to index into the disappointment behavior family. It includes behaviors that indicate sadness, which is intended to build empathy with the learner. Cosmo therefore informs the learner of the ill effect of choosing that router as he takes on a sad facial expression, slumping body language, and dropping his hands, and says, “If that were the case, we see it doesn’t arrive at the right place.”

- Rationale(Address-Resolution): To explain why the packet won't arrive at the correct destination, Cosmo adds, “This computer has no parts of the address matching.” Because the computer that serves as the referent is currently not in the focus histories and Cosmo is far from that computer, the behavior planner sequences deictic locomotion and a gesture to accompany the utterance.

- Background(Address-Resolution): The sequencing engine has selected a background speech act. Because all background and assistance speech acts cause the sequencing engine to index into the inquisitive behavior family, it obtains one of several “thoughtful” restrained manipulators such as hand waving. In this case, it selects a form of finger tapping that he performs as he explains, “Addresses are used by networked computers to tell each other apart.”

- Assistance(Address-Resolution): Finally, Cosmo assists the student by making a suggestion about the next course of action to take. Because the student has committed several mistakes on address resolution problems, Cosmo provides advice about correcting her decision by pointing to the location of the optimal computer—it has not been in focus—and stating, “This router has two parts of

the address matching."

5 Conclusions

We have discussed two characteristics of embodied conversational agents that are critical for learning environments, deictic believability and full-body emotive expression. To dynamically sequence lifelike pedagogical agents in a manner that promotes deictic believability, agent behavior planners can employ the spatial deixis framework to coordinate gesture, locomotion, and speech. To sequence full-body emotive expression, agent behavior planners can employ the emotive-kinesthetic behavior sequencing framework, which exploits the structure provided by pedagogical speech act categories to weave small emotive behaviors into larger, visually continuous ones that are responsive to students' problem-solving activities.

The spatial deixis framework and the emotive-kinesthetic framework have been informally "stress tested" in a focus group study in which ten subjects interacted with Cosmo in the Internet Protocol learning environment. Subjects unanimously expressed delight in interacting with him. Most found him fun, engaging, interesting, and charismatic. In one phase of the study in which subjects compared an "agent-free" version of the learning environment with the one inhabited by Cosmo, subjects unanimously preferred the one with Cosmo. It appeared that the learning environment with the agent clearly communicated advice with deictic speech gestures, and locomotion, though not necessarily more clearly than the agent-free version. Although some subjects voiced the opinion that Cosmo was overly dramatic, almost all exhibited particularly strong positive responses when he performed exaggerated congratulatory behaviors. In short, they found his deictic advice to be clear and helpful and his emotive to be entertaining.

This work represents a small step toward the larger goal of creating interactive, fully expressive lifelike pedagogical agents. To make significant progress in this direction, it will be important to leverage increasingly sophisticated models of human-human communication and affective reasoning. We will be pursuing these lines of investigation in our future work.

Acknowledgments

Thanks to Dorje Bellbrook, Tim Buie, Mike Cuales, Jim Dautremont, Amanda Davis, Rob Gray, Mary Hoffman, Alex Levy, Will Murray, and Roberta Osborne

of the North Carolina State University IntelliMedia Initiative for their work on the behavior sequencing engine implementation and the 3-D modeling, animation, sound, and environment design for the Internet Advisor. Thanks also to Bradford Mott for comments on an earlier draft of this chapter. Support for this work was provided by the following organizations: the National Science Foundation under grants CDA-9720395 (Learning and Intelligent Systems Initiative) and IRI-9701503 (CAREER Award Program); the North Carolina State University IntelliMedia Initiative; Novell, Inc.; and equipment donations from Apple and IBM.

References

Abou-Jaoude, S., and C. Frasson. 1998. Emotion computing in competitive learning environments. In *Working Notes of the ITS '98 Workshop on Pedagogical Agents*, 33–39.

André, E., and T. Rist. 1996. Coping with temporal constraints in multimedia presentation planning. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence*, 142–147**.** Menlo Park, Calif.: AAAI Press.

André, E., W. Finkler, W. Graf, T. Rist, A. Schauder, and W. Wahlster. 1993. WIP: The automatic synthesis of multi-modal presentations. In M. Maybury, ed., *Intelligent multimedia interfaces*, 75–93. Menlo Park, Calif.: AAAI Press.

Bates, J. 1994. The role of emotion in believable agents. *Communications of the ACM* 37(7):122–125.

Blumberg, B., and T. Galyean. 1995. Multi-level direction of autonomous creatures for real-time virtual environments. In SIGGRAPH '95, 47–54. New York: ACM.

Cassell, J., and M. Stone. 1999. Living hand to mouth: Psychological theories about speech and gesture in interactive dialogue systems. In *Proceedings of the AAAI Fall Symposium on Psychological Models of Communication in Collaborative Systems***.** Menlo Park, Calif.: AAAI Press.

Cassell, J., C. Pelachaud, N. Badler, M. Steedman, B. Achorn, T. Becket, B. Douville, S. Prevost, and M. Stone. 1994. Animated conversation: Rule-based generation of facial expression, gesture and spoken intonation for multiple conversational agents. In *SIGGRAPH'94*. New York: ACM.

Claassen, W. 1992. Generating referring expressions in a multimodal environment. In R. Dale, E. Hovy, D. Rosner, and O. Stock, eds., *Aspects of automated natural language generation*, 247–262. Berlin: Springer-Verlag.

Culhane, S. 1988. *Animation from script to screen*. New York: St. Martin's Press.

Dale, R. 1992. *Generating Referring Expressions*. Cambridge, Mass.: The MIT Press.

de Vicente, A., and H. Pain. 1998. Motivation Diagnosis in Intelligent Tutoring Systems. In *Proceedings of the Fourth International Conference on Intelligent Tutoring Systems*, 86–95. Berlin: Springer.

Elliott, C. 1992. The affective reasoner: A process model of emotions in a multi-agent system. Ph.D diss., Institute for the Learning Sciences, Northwestern University.

Elliott, C., and A. Ortony. 1992. Point of view: Reasoning about the concerns of others. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*, 809–814. Hillsdale, NJ: Lawrence Erlbaum.

Elliott, C., and G. Siegle. 1993. Variables influencing the intensity of simulated affective states. In *AAAI Spring Symposium on Reasoning about Mental States: Formal Theories and Applications*, 58–67. Menlo Park, Calif.: AAAI Press.

Elliott, C., J. Rickel, and J. Lester. 1999. Lifelike pedagogical agents and affective computing: An exploratory synthesis. In M. Wooldridge and M. Veloso, eds., *Artificial intelligence today*, 195–212. Berlin: Springer-Verlag.

Feiner, S., and K. McKeown. 1990. Coordinating text and graphics in explanation generation. In *Proceedings of the Eighth National Conference on Artificial Intelligence*, 442–449. Menlo Park, Calif.: AAAI Press.

Fillmore, C. 1975. *Santa Cruz Lectures on Deixis 1971*. Indiana University Linguistics Club. Bloomington, Ind.

Hovy, E. H. 1993. Automated discourse generation using discourse structure relations. *Artificial Intelligence* 63:341–385.

Isbister, K., and C. Nass. 1998. Personality in conversational characters: Building better digital interaction partners using knowledge about human personality preferences and perceptions. In *Notes from the Workshop on Embodied Conversational Characters*, 103–111. Tahoe City, Calif.

Jarvella, R., and W. Klein. 1982. *Speech, place, and action: Studies in deixis and related topics*. New York: John Wiley and Sons.

Johnson, W. L., and J. Rickel. 1997. Personal communication.

Johnson, W. L., J. Rickel, and J. Lester. In press. Animated pedagogical agents: Face-to-face interaction in interactive learning environments. In *The International Journal of Artificial Intelligence in Education*.

Jones, C. 1989. *Chuck amuck: The life and times of an animated cartoonist*. New York: Avon.

Kurlander, D., and D. T. Ling. 1995. Planning-based control of interface animation. In *Proceedings of CHI '95*, 472–479**.** New York: ACM Press.

Lenburg, J. 1993. *The great cartoon directors.* New York: Da Capo Press.

Lester, J. C., and B. W. Porter. 1997. Developing and empirically evaluating robust explanation generators: The Knight Experiments. *Computational Linguistics* 23(1):65–101.

Lester, J. C., S. A. Converse, S. E. Kahler, S. T. Barlow, B. A. Stone, and R. Bhogal. 1997a. The persona effect: Affective impact of animated pedagogical agents. In *Proceedings of CHI '97 Human Factors in Computing Systems*, 359–366. New York: ACM.

Lester, J. C., S. A. Converse, B. A. Stone, S. E. Kahler, and S. T. Barlow. 1997b. Animated pedagogical agents and problem-solving effectiveness: A large-scale empirical evaluation. In *Proceedings of Eighth World Conference on Artificial Intelligence in Education*, 23–30. Amsterdam: IOS Press.

Lester, J., B. Stone, and G. Stelling. 1999. Lifelike pedagogical agents for mixed-initiative problem solving in constructivist learning environments. *User Modeling and User-Adapted Interaction* 9(1–2):1–44.

Lester, J., Voerman, J., Towns, S., and Callaway, C. 1999. Deictic believability: Coordinating gesture, locomotion, and speech in lifelike pedagogical gents. *Applied Artificial Intelligence* 13(4–5):383–414.

Maybury, M. 1991. Planning multimedia explanations using communicative acts. In *Proceedings of the Ninth National Conference on Artificial Intelligence*, 61–66. Menlo Park, Calif.: AAAI Press.

Mittal, V., S. Roth, J. Moore, J. Mattis, and G. Carenini. 1995. Generating explanatory captions for information graphics. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 1276–1283. San Francisco: Morgan Kaufmann.

Moore, J. D. 1995. *Participating in explanatory dialogues*. Cambridge, Mass.: The MIT Press.

NASA. 1993. *CLIPS reference manual*. Technical report, Software Technology Branch. Lyndon B. Johnson Space Center.

Neal, J., and S. Shapiro. 1991. Intelligent multi-media interface technology. In J. Sullivan and S. Tyler, eds., *Intelligent User Interfaces*, 11–43. Reading, MA: Addison-Wesley.

Noake, R. 1988. *Animation techniques*. London: Chartwell.

Novak, H. 1987. Strategies for generating coherent descriptions of object movements in street scenes. In G. Kempen, ed., *Natural language generation*, 117–132. Dordrecht, The Netherlands: Martinus Nijhoff.

Ortony, A., G. L. Clore, and A. Collins. 1988. *The cognitive structure of emotion*. New York: Cambridge University Press.

Paiva, A., and I. Machado. 1998. Vincent, an autonomous pedagogical agent for on-the-job training. In *Proceedings of the Fourth International Conference on Intelligent Tutoring Systems*, 584–593. Berlin: Springer.

Roberts, L. 1993. *How reference works: Explanatory models for indexicals, descriptions and opacity*. New York: SUNY Press.

Roth, S., J. Mattis, and X. Mesnard. 1991. Graphics and natural language as components of automatic explanation. In J. Sullivan and S. Tyler, eds, *Intelligent user interfaces*, 207–239. Reading, MA: Addison-Wesley.

Sibun, P. 1992. Generating text without trees. *Computational Intelligence* 8(1):102–122.

Suchman, L. 1987. *Plans and situated actions: The problem of human machine communication*. New York: Cambridge University Press.

Suthers, D. 1991. A task-appropriate hybrid architecture for explanation. *Computational Intelligence* 7(4):315–333.

Thomas, F., and O. Johnston. 1981. *The illusion of life: Disney animation*. New York: Walt Disney Productions.