Serious Games Get Smart: Intelligent Game-Based Learning Environments

James C. Lester, Eun Y. Ha, Seung Y. Lee, Bradford W. Mott, Jonathan P. Rowe, Jennifer L. Sabourin

■ *Intelligent game-based learning environments integrate commercial game technologies with AI methods from intelligent tutoring systems and intelligent narrative technologies. This article introduces the Crystal Island intelligent game-based learning environment, which has been under development in the authors' laboratory for the past seven years. After presenting Crystal Island, the principal technical problems of intelligent game-based learning environments are discussed: narrative-centered tutorial planning, student affect recognition, student knowledge modeling, and student goal recognition. Solutions to these problems are illustrated with research conducted with the Crystal Island learning environment*

Recent years have seen a growing recognition of the transformative potential of game-based learning environments. The burgeoning field of game-based learning has made significant advances, including theoretical developments (Gee 2007), as well as the creation of game-based learning environments for a broad range of K–12 subjects (Habgood and Ainsworth 2011; Ketelhut et al. 2010; Warren, Dondlinger, and Barab 2008) and training objectives (Johnson 2010; Kim et al. 2009). Of particular note are the results of recent empirical studies demonstrating that in addition to game-based learning environments' potential for motivation, they can enable students to achieve learning gains in controlled laboratory settings (Habgood and Ainsworth 2011) as well as classroom settings (Ketelhut et al. 2010).

Motivated by the goal of creating game-based learning experiences that are personalized to individual students, we have been investigating intelligent game-based learning environments that leverage commercial game technologies and AI techniques from intelligent tutoring systems (D'Mello and Graesser 2010; Feng, Heffernan, and Koedinger 2009; VanLehn 2006) and intelligent narrative technologies (Lim et al. 2012; McCoy et al. 2011; Si, Marsella, and Pynadath 2009; Yu and Riedl 2012). This work ranges from research on real-time narrative planning and goal recognition to affective computing models for recognizing student emotional states. Intelligent game-based learning environments serve as an excellent laboratory for investigating AI techniques because they make significant inferential demands and play out in complex interactive scenarios.

In this article we introduce intelligent game-based learning

 ISSN 0738-4602

Figure 1. Crystal Island Narrative-Centered Learning Environment.

environments by presenting Crystal Island, an intelligent game-based learning environment for middle grade science education (Rowe et al. 2011). Crystal Island has been under continual development through a series of learning technology investigations and laboratory and classroom studies over the past seven years. We discuss technical problems that we have investigated in the context of Crystal Island, including narrative-centered tutorial planning, student affect recognition, student modeling, and student goal recognition. We conclude with a discussion of educational impacts of intelligent game-based learning environments, and future directions for the field.

The Crystal Island Learning Environment

Crystal Island (figure 1) is a narrative-centered learning environment that was originally built on Valve Software's Source engine, the three-dimensional (3-D) game platform for Half-Life 2, and now runs on the Unity cross-platform game engine from Unity Technologies. The curriculum underlying Crystal Island's mystery narrative is derived from the North Carolina state standard course of study for eighth-grade microbiology. The environment is designed as a supplement to classroom instruction, and it blends elements of both inquiry learning and direct instruction. Crystal Island has served as a platform for investigating a range of artificial intelligence technologies for dynamically supporting students' learning experiences. This includes work on narrative-centered tutorial planning (Lee, Mott, and Lester 2012; Mott and Lester 2006), student knowledge modeling (Rowe and Lester 2010), student goal recognition (Ha et al. 2011), and affect recognition models (Sabourin, Mott, and Lester 2011). The environment has also been the subject of extensive empirical investigations of student learning and presence (Rowe et al. 2011), with results informing the design and revision of successive iterations of the system.

Crystal Island features a science mystery where students attempt to discover the identity and source of an infectious disease that is plaguing a research team stationed on a remote island. Students adopt the role of a medical field agent who has been assigned to investigate the illness and save the research team from the outbreak. Students explore the research camp from a first-person viewpoint and manipulate virtual objects, converse with characters, and use lab equipment and other resources to solve the mystery. The mystery is solved when students complete a series of partially ordered goals that involve uncovering details about the spreading infection, testing potential transmission sources of the disease, recording a diagnosis and treatment plan, and presenting the findings to the camp nurse.

The following scenario illustrates a typical interaction with Crystal Island. The scenario begins with the

Figure 2. Camp Nurse and Sick Patient in Virtual Infirmary.

player-character disembarking from her boat shortly after arriving on the island. After completing a brief tutorial that introduces the game's controls, the player leaves the dock to approach the research camp. Near the camp entrance, the student encounters an infirmary with several sick patients and a camp nurse. Upon entering the infirmary, the student approaches the nurse and initiates a conversation (figure 2). The nurse explains that an unidentified illness is spreading through the camp and asks for the player's help in determining a diagnosis. She advises the student to use an in-game diagnosis worksheet in order to record her findings, hypotheses, and final diagnosis (figure 3). This worksheet is designed to scaffold the student's problem-solving process, as well as provide a space for the student to offload any findings gathered about the illness. The conversation with the nurse takes place through a combination of multimodal character dialogue — spoken language, gesture, facial expression, and text — and player dialogue menu selections.

After speaking with the nurse, the student has several options for investigating the illness. The student can talk to sick patients lying on medical cots in order to gather information about the team members' symptoms and recent eating habits. Alternatively, the student can move to the camp's dining hall to speak with the camp cook. The cook describes the types of food that the team has recently eaten and provides clues about which items warrant closer investigation. In addition to learning about the sick team members, the student can walk to the camp's living quarters to converse with a pair of virtual scientists who answer questions about viruses and bacteria. The student can also learn more about pathogens by viewing posters hanging inside of the camp's buildings or reading books located in a small library.

Beyond gathering information about the disease from virtual scientists and other instructional resources, the student can test potential transmission sources using the laboratory's testing equipment. For example, the student encounters several food items that have been lying out in the dining hall, and she can test the items for infectious agents at any point during the learning interaction.

After running several tests, the student discovers that the sick team members have consumed milk that is contaminated with a bacterial agent. The student can use the camp's books and posters in order to investigate bacterial diseases that are associated with symptoms matching those reported by the sick team members. Once she has narrowed down a diagnosis and recommended treatment, the student returns to the infirmary in order to inform the camp nurse. If the student's diagnosis is incorrect, the nurse identifies the error and recommends that the player keep working. The student can use this feedback to roughly determine how close she is to solving the mystery. When the student correctly diagnoses the illness and specifies an appropriate treatment, the mystery is solved.

Figure 3. Diagnosis Worksheet Where Students Record their Findings.

Decision-Theoretic Narrative Planning

In order to perform real-time adaptive personalization of interactive narratives in game-based learning environments like Crystal Island, director agents (or drama managers) plan unfolding narratives by operating on at least two distinct but interacting levels: they craft the global story arc, typically by traversing a plot graph that encodes a partial order of significant events in the narrative, and they plan virtual character behaviors and physical events in the storyworld. This task, known as narrative-centered tutorial planning, blends aspects of interactive narrative planning and tutorial planning in a single problem. Interactive narrative planning encompasses a number of challenges, such as balancing character believability and plot coherence in the presence of unpredictable user actions (Riedl and Young 2004), or providing rich character behaviors that are consistent with authorial objectives (Mateas and Stern 2005). The task's tutorial planning component requires that the director agent guide students' cognitive, affective, and metacognitive processes to promote effective learning outcomes.

To address these requirements for Crystal Island, we designed and implemented U-Director (Mott and Lester 2006), a decision-theoretic narrative planning architecture that uses a dynamic decision network (DDN) (Dean and Kanazawa 1989) to model narrative objectives, storyworld state, and user state in order to inform run-time adaptations of Crystal Island's science mystery. A DDN-based approach to narrative planning provides a principled decision-making framework for reasoning about the multiple sources of evidence that inform narrative-centered tutorial planning. At each game clock cycle, U-Director systematically evaluates available evidence, updates its beliefs, and selects the storyworld action that maximizes expected narrative-tutorial utility.

Because narrative is fundamentally a time-based phenomenon, in each decision cycle U-Director considers candidate narrative actions to project forward in time the effects of the actions being taken and their consequent effects on the user (Mott and Lester 2006). To do so, it evaluates its narrative objectives in light of the current storyworld state and user state. Each decision cycle considers three distinct time slices (*narrative state*$_t$, *narrative state*$_{t+1}$, and *narrative state*$_{t+2}$), each of which consists of interconnected subnetworks containing chance nodes in the DDN (figure 4). The three slices represent (1) the current narrative state, (2) the narrative state after the director agent's decision, and (3) the narrative state after the user's next action. The DDN's director action is a decision node, the DDN's user action is a chance node, and *utility*$_{t+2}$ is a utility node in the DDN. Each time slice encodes a probabilistic representation of

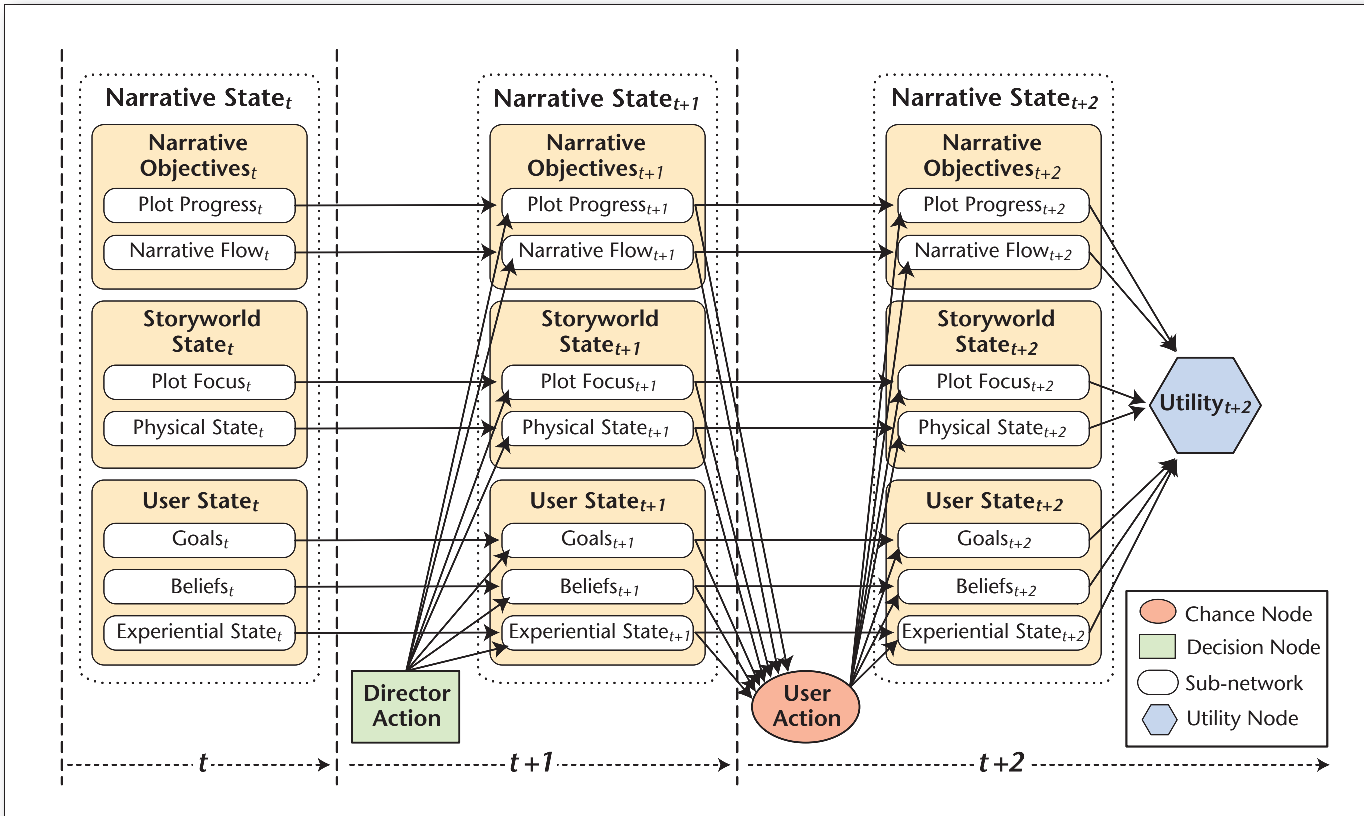

Figure 4. High-Level Illustration of Dynamic Decision Network for Interactive Narrative Director Agent.

the director's beliefs about the overall state of the narrative, represented with narrative-centered knowledge sources.

U-Director begins its computation by considering narrative state$_t$, which represents the model's current beliefs about the unfolding story's narrative objectives, storyworld state, and user state (Mott and Lester 2006). Links from the director action node to the *narrative state*$_{t+1}$ node model candidate director actions and how they affect the story. Example director actions include instructing nonplayer characters to perform actions in the environment or guiding students through the story by providing appropriately leveled hints. U-Director constrains the number of candidate director actions to evaluate at each time step using abstract director actions, thereby limiting the number of concrete actions considered. Next, it models how the possible worlds encoded in *narrative state*$_{t+1}$ influence the user's action and how the user's action in turn affects the story in *narrative state*$_{t+2}$. Finally, using links from *narrative state*$_{t+2}$ to the utility node *utility*$_{t+2}$, U-Director models preferences over potential narrative states. Preferences provide a representation in which authors specify the relative importance of salient features of the narrative state.

To gauge the overall effectiveness of U-Director's narrative planning capabilities a simulated user approach was taken (Mott and Lester 2006). Six simulated users (three cooperative users who typically followed the agent's guidance and three uncooperative users who typically did not follow the agent's guidance) interacted with the Crystal Island director agent to create six different narrative experiences. Traces of the director agent's decision making were analyzed to see if the agent took appropriate action to guide users through the narrative. As was desired, the analysis revealed that the director agent adopted a hint-centered approach for cooperative users and a more heavy-handed approach for uncooperative users. The simulated user approach offers a promising means for establishing baseline performance prior to conducting extensive focus group studies with human users.

In addition to formalisms based on dynamic decision networks, we have also experimented with empirically based models of narrative-centered tutorial planning using a Wizard-of-Oz framework, where human wizards serve as expert narrative-centered tutorial planners (Lee, Mott, and Lester 2012). In this approach, trained wizards interact with students who are attempting to solve Crystal Island's science mystery. The wizards perform tutorial and narrative planning functionalities by controlling a virtual character and guiding students through the game-based learning environment. Detailed trace data from these wizard-student interactions is collected by the virtual environment — including all wizard decision-making, navigation, and manipulation activities — in order to generate a training corpus for inducing narrative-centered tutorial planning models using supervised machine-learning techniques.

To explore the run-time effectiveness of the machine-learned narrative-centered tutorial planning models, induced intervention and action models were integrated into Crystal Island, where the models determine when is the most appropriate time to intervene and what is the most appropriate next narrative-tutorial action. This extended version of Crystal Island was identical to the Wizard-of-Oz-based version, except a narrative-centered tutorial planner rather than a human wizard drove nonplayer character interactions. The planning models actively observe students' activities and dynamically guide students by directing characters in the virtual storyworld.

Three experimental conditions were created to evaluate the effectiveness of the induced narrative-centered tutorial planner: minimal guidance, intermediate guidance, and full guidance. Outcomes of the three conditions were compared to determine the effectiveness of utilizing our machine-learned models.

Minimal Guidance: Students experience the storyworld under the guidance of a minimal narrative-centered tutorial planning model. This model controls events that are required to be performed by the system (that is, students cannot complete the events without the system taking action). The model in this condition is not machine learned. It performs an action once all preconditions are met for the action to be taken. This condition served as a baseline for the investigation.

Intermediate Guidance: Students experience the storyworld under the guidance of an intermediate narrative-centered tutorial planning model. This is an ablated model inspired by the notions of interactive narrative islands (Riedl et al. 2008). Islands are intermediate plan steps through which all valid solution paths must pass. They have preconditions describing the intermediate world state, and if the plan does not satisfy each island's preconditions, the plan will never achieve its goal. In our version of Crystal Island, transitions between narrative arc phases represent "islands" in the narrative. Each arc phase consists of a number of potential narrative-centered tutorial planning decisions. However, the phases are bounded by specific narrative-centered tutorial planning decisions that define when each phase starts and ends. We employ these specific tutorial action decisions as our islands.

Full Guidance: Students experience the storyworld under the guidance of the full narrative-centered tutorial planning model. The model actively monitors students interacting with the storyworld in order to determine when it is appropriate to intervene with the next tutorial action. The model has full control of the tutorial intervention decisions (that is, determining when to intervene) and tutorial action decisions (that is, determining what the intervention should be). The full guidance tutorial planning model employs the entire set of narrative-centered tutorial planning decisions available to it. This list is described in Lee, Mott, and Lester (2011).

In an experiment comparing the three narrative-centered tutorial planning models, a total of 150 eighth-grade students used Crystal Island and completed the pre- and posttest measures (Lee, Mott, and Lester 2012). The students were drawn from a suburban middle school. While the study was held during the school day, groups of students were pulled out of class to play the game in a laboratory setting. The pre- and posttests consisted of the same multiple-choice questions about Crystal Island's microbiology curriculum. Students completed the pretest several days prior to using Crystal Island, and they completed the posttest immediately after solving the science mystery. In all three conditions, the tutorial planning models' actions focused on introducing narrative events, or providing problem-solving advice to students. The tutorial planner did not perform actions involving direct instruction of curricular content.

An investigation of overall learning found that students who interacted with Crystal Island achieved positive learning gains on the curriculum test. A two-tailed matched pairs t-test between posttest and pretest scores indicates that the learning gains were significant, $t(149) = 2.03$, $p < .05$. It was also found that students' Crystal Island interactions in the full guidance condition yielded significant learning gains, as measured by the difference of posttest and pretest scores. As shown in table 1, a two-tailed matched pairs t-test showed that students in the full guidance condition showed statistically significant learning gains. Students in the intermediate and minimal guidance conditions did not achieve significant learning gains.

In addition, we analyzed learning gain differences between the conditions and the results showed that there were significant differences. Performing an ANCOVA to control for pretest scores, the learning gains were significantly different for the full and minimal guidance conditions, $F(2, 99) = 38.64$, $p < .001$, as were the learning gains for the full and intermediate guidance conditions, $F(2, 100) = 40.22$, $p < .001$. Thus, students who received full guidance from our machine-learned models achieved significantly higher learning gains than the students who were in the other two conditions. These results are consistent with findings from the education literature, which suggest that students who receive problem-solving guidance (coaching, hints) during inquiry-based learning achieve greater learning outcomes than students who receive minimal or no guidance (Kirschner, Sweller, and Clark 2006; Mayer 2004).

Modeling Student Affect

Affect has begun to play an increasingly important role in game-based learning environments. The intelligent tutoring systems community has seen the emergence of work on affective student modeling (Conati and Maclaren 2009), including models for detecting frustration and stress (McQuiggan, Lee, and Lester 2007), modeling agents' emotional states (Marsella and Gratch 2009), and detecting student motivation (de Vicente and Pain 2002). All of this work seeks to increase the fidelity with which affective and motivational processes are understood and utilized in intelligent tutoring systems in an effort to increase the effectiveness of tutorial interactions and, ultimately, learning.

This focus on examining affect is largely due to the effects it has been shown to have on learning outcomes. Emotional experiences during learning have been shown to affect problem-solving strategies, the level of engagement exhibited by students, and the degree to which the student is motivated to continue with the learning process (Picard et al. 2004). These factors have the power to dictate how students learn immediately and their learning behaviors in the future. Consequently, the ability to understand and model affective behaviors in learning environments has been a focus of recent work (Arroyo et al. 2009; Conati and Maclaren 2009; D'Mello and Graesser 2010).

Correct prediction of students' affective states is an important first step in designing affect-sensitive game-based learning environments. The delivery of appropriate affective feedback requires first that the student's state be accurately identified. However, the detection and modeling of affective behaviors in learning environments poses significant challenges. For example, many successful approaches to affect detection utilize a variety of physical sensors to inform model predictions. However, deploying these types of sensors in schools or home settings often proves to be difficult, or impractical, due to concerns about cost, privacy, and invasiveness. Consequently, reliance on physical sensors when building affect-sensitive learning environments can limit the potential scalability of these systems. To combat this issue, many systems attempt to model emotion without the use of physiological sensors. Some researchers taking this approach have focused on incorporating theoretical models of emotion, such as appraisal theory, which is particularly well suited for computational environments (Marsella and Gratch 2009). These models propose that individuals appraise events and actions in their environment according to specific criteria (for example, desirability or cause) to arrive at emotional experiences. While there are a variety of appraisal-based theories of emotions, few appraisal models focus specifically on the emotions that typically occur during learning (Picard et al. 2004).

Conditions	Students	Gain Avg.	SD	t	p
Full	55	1.28	2.66	2.03	< 0.05
Intermediate	48	0.13	2.69	0.19	0.84
Minimal	47	0.89	3.12	1.23	0.22

Table 1. Learning Gain Statistics for Comparison of Induced Narrative-Centered Tutorial Planning Models.

The lack of a widely accepted and validated model of learner emotions poses a challenge for the development of affect-detection systems using theoretical grounding in place of physical sensors. However, Elliot and Pekrun's model of learner emotions describes how students' affective states relate to their general goals during learning tasks, such as whether they are focused on learning or performance, and how well these goals are being met (Elliot and Pekrun 2007). The generality of this model allows it to be adapted to specific learning tasks and makes it well suited for predictive modeling in open-ended learning environments.

Elliot and Pekrun's model of learner emotions was used as a theoretical foundation for structuring a sensor-free affect detection model (Sabourin, Mott, and Lester 2011). This model was empirically learned from a corpus of student interaction data. During their interactions with Crystal Island, students were asked to self-report on their affective state through an in-game smartphone device. They were prompted every seven minutes to select one emotion from a set of seven options, which included anxious, bored, confused, curious, excited, focused, and frustrated. This set of cognitive-affective states is based on prior research identifying states that are relevant to learning (Craig et al. 2004, Elliot and Pekrun 2007). Each emotional state was also accompanied by a validated emoticon to provide clarity to the emotion label.

A static Bayesian network (figure 5) was designed with the structure hand-crafted to include the relationships described within Elliot and Pekrun's model of learner emotions (Sabourin, Mott, and Lester 2011). Specifically the structure focused on representing the appraisal of learning and performance goals and how these goals were being met based on students' progress and activities in the game. For example, some activities such as book reading or note taking in Crystal Island are related to learning objectives, while achievement of certain milestones and external validation are more likely to contribute to measures of performance goals. Several other components of Elliot and Pekrun's model are also considered. For example, students with approach orientations are expected to have generally more positive temperaments and emotional experiences than students with avoidance orientations.

Using a training corpus of student interaction

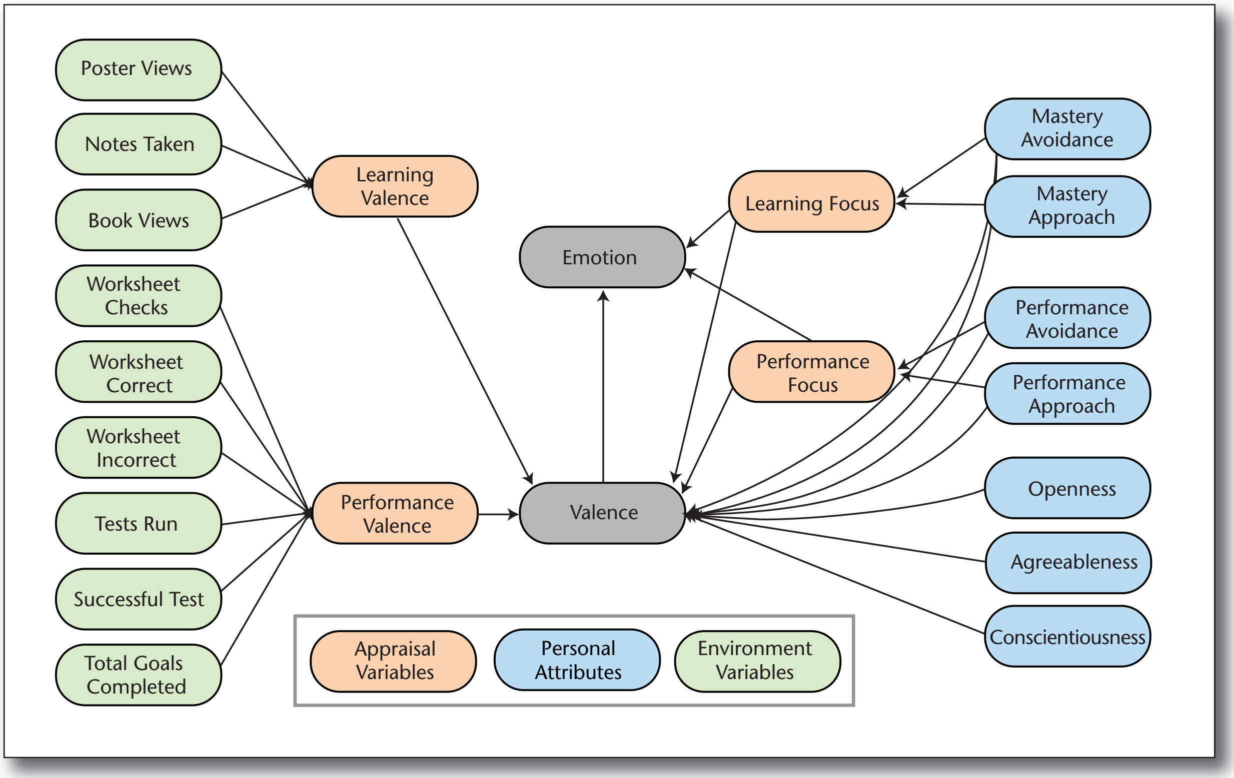

Figure 5. Structure of Static Bayesian Network Model of Learner Affect.

data to learn and validate the parameters of static and dynamic Bayesian network models (Sabourin, Mott, and Lester 2011), it was found that the Elliot and Pekrun (2007) model of learner emotions can successfully serve as the basis for computational models of learner emotions. The model grounded by a theoretical understanding of learner emotions outperformed several baseline measures including most-frequent class as well as a Bayesian network model operating under naïve variable independence assumptions (table 2). It was also found that modeling the dynamic quality of emotional states across time through a dynamic Bayesian network offered significant improvement in performance over the static Bayesian network. The model performed particularly well at identifying positive states, including the state of focused, which was the most frequent emotion label (tables 3–4). The model had difficulty distinguishing between the states of confusion and frustration, which is intriguing because the causes and experiences of these two states are often very similar. These findings indicate the need for future work to improve predictions of negative emotional states so that models for automatic detection of students' learning emotions can be used for affective support in game-based learning environments.

Student Modeling with Dynamic Bayesian Networks

Devising effective models of student knowledge in game-based learning environments poses significant computational challenges. First, models of student knowledge must cope with multiple sources of uncertainty inherent in the modeling task. Second, knowledge models must dynamically model knowledge states that change over the course of a narrative interaction. Third, the models must concisely represent complex interdependencies among different types of knowledge, and naturally incorporate multiple sources of evidence about user knowledge. Fourth, the models must operate under the real-time performance constraints of the interactive narratives that play out in intelligent game-based learning environments.

To address these challenges, we developed a dynamic Bayesian network (DBN)) approach to modeling user knowledge during interactive narrative experiences (Rowe and Lester 2010). DBNs offer a unified formalism for representing temporal stochastic processes such as those associated with knowledge modeling in interactive narrative environments. The framework provides a mechanism for dynamically updating a set of probabilistic beliefs about a student's understanding of narrative, mystery solution, strategic, and curricular knowledge components that are accumulated and demonstrated during interactions with a narrative environment. An initial version of the model has been implemented in Crystal Island.

The DBN for knowledge tracing was implemented with the SMILE Bayesian modeling and inference library developed by the University of Pittsburgh's Decision Systems Laboratory (Druzdzel 1999). The model maintains approximately 135 binary nodes, 100 directed links, and more than 750 conditional probabilities. As the knowledge-tracing model observes student actions in the environment, the associated evidence is incorporated into the network, and a Bayesian update procedure is performed. The update procedure, in combination with the network's singly connected structure, yields updates that complete in less than one second. Initial probability values were fixed across all students; probabilities were chosen to represent the assumption that students at the onset had no understanding of scenario-specific knowledge components and were unlikely to have mastery of curriculum knowledge components.

A human participant study was conducted with 116 eighth-grade students from a middle school interacting with the Crystal Island environment (Rowe and Lester 2010). During the study, students interacted with Crystal Island for approximately 50 minutes. Logs of students' in-game actions were recorded, and were subsequently used to conduct an evaluation of the DBN knowledge-tracing model. The evaluation aimed to assess the model's ability to accurately predict students' performance on a content knowledge postexperiment test. While the evaluation does not inspect the student knowledge model's intermediate states during the narrative interaction, nor students' narrative-specific knowledge, an evaluation of the model's final knowledge assessment accuracy provided a useful starting point for refining the model and evaluating its accuracy.

The DBN knowledge-tracing model used the students' recorded trace data as evidence to approximate students' knowledge at the end of the learning interaction. This yielded a set of probability values for each student, corresponding to each of the knowledge-tracing model's knowledge components. The resultant data was used in the analysis of the model's ability to accurately predict student responses on postexperiment content test questions. The mapping between the model's knowledge components and individual posttest questions was generated by a researcher, and used the following heuristic: if a posttest question or correct response shared important content terms with the description of a particular knowledge component, that knowledge component was designated as necessary for providing an informed, correct response to the question. According to this heuristic, several questions required the simultaneous application of multiple knowledge components, and a number of knowledge components bore on multiple questions. This yielded a many-to-many mapping between knowledge components and posttest questions.

The evaluation procedure required the definition of a threshold value to discriminate between mastered and unmastered knowledge components: knowledge components whose model values exceeded the threshold were considered mastered, and knowledge components whose model values fell below the threshold were considered unmastered. The model predicted a correct response on a posttest question if all of the question's associated knowledge components were considered mastered. The model predicted an incorrect response on a posttest question if one or more associated knowledge components were considered unmastered. The use of a threshold to discriminate between mastered and unmastered knowledge components mirrors how the knowledge model might be used in a run-time

	Emotion Accuracy	Valence Accuracy
Baseline	22.4%	54.5%
Naïve Bayes	18.1%	51.2%
Bayes Net	25.5%	66.8%
Dynamic BN	32.6%	72.6%

Table 2. Prediction Accuracies.

Actual Emotion	Correct Emotion Prediction	Correct Valence Prediction
anxious	2%	60%
bored	18%	75%
confused	32%	59%
curious	38%	85%
excited	19%	79%
focused	52%	81%
frustrated	28%	56%

Table 3. Prediction Accuracy by Emotion.

		Predicted Valence	
		Positive	Negative
Actual Valence	Positive	823	184
	Negative	326	512

Table 4. Valence Confusion Matrix.

environment to inform interactive narrative decision making.

Rather than choose a single threshold, a series of values ranging between 0.0 and 1.0 were selected. For each threshold, the DBN knowledge model was compared to a random model, which assigned uniformly distributed, random probabilities for each knowledge component. New random probabilities were generated for each knowledge component, student, and threshold. Both the DBN model and random model were used to predict student posttest responses, and accuracies for each threshold were determined across the entire test. Accuracy was measured as the sum of successfully predicted correct responses plus the number of successfully predicted incorrect responses, divided by the total number of questions.

The DBN model outperformed a random baseline model across a range of threshold values. The DBN model most accurately predicted students' posttest responses at a threshold level of 0.32 (M = .594, SD = .152). A Wilcoxon-Mann-Whitney U test verified that the DBN knowledge-tracing model was significantly more accurate than the random model at the 0.32 threshold level, $z = 4.79$, $p < .0001$. Additional Mann-Whitney tests revealed that the DBN model's predictive accuracy was significantly greater than that of the random model, at the $\alpha = .05$ level, for the entire range of thresholds between .08 and .56.

Student Goal Recognition

Goal recognition, as well as its sibling tasks, plan recognition and activity recognition, are long-standing AI problems (Charniak and Goldman 1993; Kautz and Allen 1986; Singla and Mooney 2011) that are central to game-based learning. The problems are cases of abduction: given domain knowledge and a sequence of actions performed by an agent, the task is to infer which plan or goal the agent is pursuing. Recent work has yielded notable advances in recognizing agents' goals and plans, including methods for recognizing multiple concurrent and interleaved goals (Hu and Yang 2008), methods for recognizing activities in multiagent settings (Sadilek and Kautz 2010), and methods for augmenting statistical relational learning techniques to better support abduction (Singla and Mooney 2011).

Digital games pose significant computational challenges for goal-recognition models. For example, in many games players' abilities change over time, both by unlocking new powers and improving motor skills over the course of game play. In effect, a player's action model may change, in turn modifying the relationships between actions and goals. Action failure is also a critical and deliberate design choice in games; in platform games, a poorly timed jump often leads to a player's demise. In multiplayer games, multiagent goals arise that may involve players competing or collaborating to accomplish game objectives. Individual players may also pursue ill-defined goals, such as "explore" or "try to break the game." In these cases goals and actions may be cyclically related; players take actions in pursuit of goals, but they may also choose goals after they are revealed by particular actions. For these reasons, game-based learning environments offer promising test beds for investigating different formulations of goal-recognition tasks.

To investigate these issues, we have been devising goal-recognition models for Crystal Island (Ha et al. 2011). Given its relationship to abduction, goal recognition appears well suited for logical representation and inference. However, goal recognition in digital games also involves inherent uncertainty. For example, a single sequence of actions is often explainable by multiple possible goals. Markov logic networks (MLNs) provide a formalism that unifies logical and probabilistic representations into a single framework (Richardson and Domingos 2006). To address the problem of goal recognition with exploratory goals in game environments, a Markov logic goal-recognition framework was devised.

The MLN goal-recognition model was trained on a corpus collected from player interactions with Crystal Island (Ha et al. 2011). In this setting, goal recognition involves predicting the next narrative subgoal that the student will complete as part of investigating the mystery.

An MLN consists of a set of weighted first-order logic formulae. A weight reflects the importance of the constraint represented by its associated logic formula in a given model. Figure 6 shows 13 MLN formulae that are included in the goal-recognition model. Formula 1 represents a hard constraint that needs to be satisfied at all times. This formula requires that, for each action a at each time step t, there exists a single goal g. The formulae 2–13 are soft constraints that are allowed to be violated. Formula 2 reflects the prior distribution of goals in the corpus. Formulae 3–13 predict the player's goal g at time t based on the values of the three action properties, action type a, location l, and narrative state s, as well as the previous goal. The weights for the soft formulae were learned with theBeast, an off-the-shelf tool for MLNs that uses cutting plane inference (Riedel 2008).

Similar to the U-Director framework described above, the MLN framework encodes player actions in the game environment using three properties: action type, location, and narrative state.

Action Type: Type of action taken by the player, such as moving to a particular location, opening a door, and testing an object using the laboratory's testing equipment. Our data includes 19 distinct types of player actions.

Location: Place in the virtual environment where a player action is taken. This includes 39 fine-grained and nonoverlapping sublocations that decompose the seven major camp locations in Crystal Island.

Narrative State: Representation of the player's

Hard Formula

$$\forall t,a : action(t,a) \Rightarrow |\exists g : goal(t,g)| = 1 \quad (1)$$

Soft Formulae

$$\forall t,g : goal(t,g) \quad (2)$$
$$\forall t,a,g : action(t,a) \Rightarrow goal(t,g) \quad (3)$$
$$\forall t,l,g : loc(t,l) \Rightarrow goal(t,g) \quad (4)$$
$$\forall t,s,g : state(t,s) \Rightarrow goal(t,g) \quad (5)$$
$$\forall t,a,s,g : action(t,a) \wedge state(t,s) \Rightarrow goal(t,g) \quad (6)$$
$$\forall t,a,g : action(t-1,a) \Rightarrow goal(t,g) \quad (7)$$
$$\forall t,l,g : loc(t-1,l) \Rightarrow goal(t,g) \quad (8)$$
$$\forall t,s,g : state(t-1,s) \Rightarrow goal(t,g) \quad (9)$$
$$\forall t,a,s,g : action(t-1,a) \wedge state(t-1,s) \Rightarrow goal(t,g) \quad (10)$$
$$\forall t,a_1,a_2,g : action(t-1,a_1) \wedge action(t,a_2) \Rightarrow goal(t,g) \quad (11)$$
$$\forall t,g_1,g_2 : goal(t-1,g_1) \Rightarrow goal(t,g_2) \quad (12)$$
$$\forall t,a_1,a_2,g_1,g_2 : action(t-1,a_1) \wedge goal(t-1,g_1) \wedge action(t,a_2) \Rightarrow goal(t,g_2) \quad (13)$$

Figure 6. Formulae for MLN Goal-Recognition Model.

progress in solving the narrative scenario. Narrative state is encoded as a vector of four binary variables, each of which represents a milestone event within the narrative.

The performance of the proposed MLN goal-recognition model was compared to one trivial and two nontrivial baseline models (Ha et al. 2011). The trivial baseline was the majority baseline, which always predicted the goal that appears most frequently in the training data. The nontrivial baselines were two *n*-gram models, unigram and bigram. The unigram model predicted goals based on the current player action only, while the bigram model considered the previous action as well. In the *n*-gram models, player actions were represented by a single aggregate feature that combined all three action properties: action type, location, and narrative state. In addition to game environments, *n*-gram models have been used in previous goal-recognition work for spoken dialogue systems (Blaylock and Allen 2003).

The two *n*-gram models and the proposed MLN model were evaluated with tenfold cross validation. The entire data set was partitioned into ten nonoverlapping subsets, ensuring data from the same player did not appear in both the training and the testing data. Each subset of the data was used for testing exactly once. The models' performance was measured using F1, which is the harmonic mean of precision and recall. Table 5 shows the average performance of each model over ten-fold cross validation. The MLN model scored 0.484 in F1, achieving an 82 percent improvement over the baseline. The unigram model performed better than the bigram model. A one-way repeated-measures ANOVA confirmed that the differences among the three compared models were statistically significant ($F(2, 18) = 71.87$, $p < 0.0001$). A post hoc Tukey test revealed the differences between all pairs of the three models were statistically significant ($p < .01$).

Educational Impact

In addition to the version of Crystal Island for middle school science education described above, two additional versions of Crystal Island have been developed: one for elementary science education, and one for integrated science and literacy education for middle school students. More than 4000 students have been involved with studies with the Crystal Island learning environments. While the learning

	Baseline	Unigram	Bigram	Factored MLN
F1	0.266	0.396	0.330	0.484
Improvement over Baseline	N/A	49%	24%	82%

Table 5. F1 Scores for MLN and Baseline Goal-Recognition Models.

Predictor	B	SE	β
Pretest	.46**	.09	.33**
Presence	.03*	.01	.15*
Final Game Score	.01**	.00	.31**
$R^{`1}$		.33	

Table 6. Regression Results Predicting Students' Microbiology Posttest Performance.

Note: ** - $p < .01$; * - $p < .05$

environments are in continual development, they regularly achieve learning gains in students. For example, an observational study involving 150 students investigated the relationship between learning and engagement in Crystal Island (Rowe et al. 2011). All students in the study played Crystal Island. The investigation explored questions in the science education community about whether learning effectiveness and engagement are synergistic or conflicting in game-based learning. Students were given pre- and posttests on the science subject matter. An investigation of learning gains found that students answered more questions correctly on the posttest ($M = 8.61$, $SD = 2.98$) than the pretest ($M = 6.34$, $SD = 2.02$), and this finding was statistically significant, $t(149) = 10.49$, $p < .001$.

The relationship between learning and engagement was investigated by analyzing students' learning gains, problem-solving performance, and several engagement-related factors. The engagement-related factors included presence, situational interest, and in-game score. Rather than finding an oppositional relationship between learning and engagement, the study found a strong positive relationship between learning outcomes, in-game problem solving, and increased engagement (table 6). Specifically, a linear regression analysis found that microbiology background knowledge, presence, and in-game score were significant predictors of microbiology posttest score, and the model as a whole was significant, $R^2 = .33$, $F(3, 143) = 23.46$, $p < .001$. A related linear regression analysis found that microbiology pretest score, number of in-game goals completed, and presence were all significant predictors of microbiology posttest performance, $R^2 = .35$, $F(4, 127) = 16.9$, $p < .01$. Similar results have been found with the elementary school version of Crystal Island. With 800 students over twelve 50-minute class periods, students' science content test scores increased significantly, the equivalent of a letter grade.

Related Work

Intelligent game-based learning environments leverage advances from two communities: intelligent tutoring systems and intelligent narrative technologies. Intelligent tutoring systems model key aspects of one-on-one human tutoring in order to create learning experiences that are individually tailored to students based on their cognitive, affective, and metacognitive states (Woolf 2008). By maintaining explicit representations of learners' knowledge and problem-solving skills, intelligent tutoring systems can dynamically customize problems, feedback, and hints to individual learners (Koedinger et al. 1997, VanLehn 2006). Recent advances in intelligent tutoring systems include improvements to fine-grained, temporal models of student knowledge acquisition (Baker, Goldstein, and Heffernan 2011); models of tutorial dialogue strategies that enhance students' cognitive and affective learning outcomes (Forbes-Riley and Litman 2011); models of students' affective states and transitions during learning (Conati and Maclaren 2009, D'Mello and Graesser 2010); machine-learning-based techniques for embedded assessment (Feng, Heffernan, and Koedinger 2009); and tutors that model and directly enhance students' self-regulated learning skills (Azevedo et al. 2010, Biswas et al. 2010). Critically, the field has converged on a set of scaffolding functionalities that yield improved student learning outcomes compared to nonadaptive techniques (VanLehn 2006).

Intelligent narrative technologies model human story telling and comprehension processes, including methods for generating interactive narrative experiences that develop and adapt in real time. Given the central role of narratives in human cognition and communication, there has been growing interest in leveraging computational models of narrative for a broad range of applications in training (Johnson 2010; Kim et al. 2009; Si, Marsella, and Pynadath 2009) and entertainment (Lin and Walker 2011; McCoy et al. 2011; Porteous et al. 2011; Swanson and Gordon 2008; Yu and Riedl 2012). In educational settings, computational models of interactive narrative can serve as the basis for story-centric problem-solving scenarios that adapt story-centered instruction to individual learners (Johnson 2010; Lim et al. 2012; Rowe et al. 2011).

Educational research on story-centered game-based learning environments is still in its nascent stages. While there have been several examples of successful story-centric game-based learning environments for K–12 education (Ketelhut et al. 2010; Lim et al. 2012; Warren, Dondlinger, and Barab 2008) and training applications (Johnson 2010), considerable work remains to establish an empirical account of which educational settings and game features are best suited for promoting effective learning and high engagement. For example, recent experiments conducted with college psychology students have indicated that narrative-centered learning environments may not always enhance short-term content retention more effectively than direct instruction (Adams et al. 2012). However, the instructional effectiveness of story-centered game-based learning environments is likely affected by their instructional designs, game designs, and capacity to tailor scaffolding to individual learners using narrative-centered tutorial planning, student affect recognition, student modeling, and student goal recognition. Additional empirical research is needed to determine which features, and artificial intelligence techniques, are most effective in various settings. Furthermore, intelligent game-based learning environments have a broad range of applications for guided inquiry-based learning, preparation for future learning, assessment, and learning in informal settings such as museums and homes. Identifying how integrated AI systems that combine techniques from intelligent tutoring systems and intelligent narrative technologies can best be utilized to enhance robust, lifelong learning is a critical challenge for the field.

Conclusion

Intelligent game-based learning environments artfully integrate commercial game technologies and AI frameworks from intelligent tutoring systems and intelligent narrative technologies to create personalized learning experiences that are both effective and engaging. Because of their ability to dynamically tailor narrative-centered problem-solving scenarios to customize advice to students, and to provide real-time assessment, intelligent game-based learning environments offer significant potential for learning both in and out of the classroom.

Over the next few years as student modeling and affect recognition capabilities become even more powerful, intelligent game-based learning environments will continue to make their way into an increasingly broad range of educational settings spanning classrooms, homes, science centers, and museums. With their growing ubiquity, they will be scaled to new curricula and serve students of all ages. It will thus become increasingly important to understand how students can most effectively interact with them, and what role AI-based narrative and game mechanics can play in scaffolding learning and realizing sustained engagement. Further, continued improvements in the accuracy, scalability, robustness, and ease of authorship of AI models for narrative-centered tutorial planning, student affect recognition, student modeling, and student goal recognition will be critical for actualizing wide-scale use of intelligent game-based learning environments in learners' everyday lives.

Acknowledgements

The authors wish to thank members of the IntelliMedia Group of North Carolina State University for their assistance. This research was supported by the National Science Foundation under Grants REC-0632450, DRL-0822200, IIS-0812291, and CNS-0540523. This work was also supported by the National Science Foundation through a Graduate Research Fellowship. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. This work was also supported by the Bill and Melinda Gates Foundation, the William and Flora Hewlett Foundation, and Educause.

References

Adams, D. M.; Mayer, R. E.; Macnamara, A.; Koenig, A.; and Wainess, R. 2012. Narrative Games for Learning: Testing the Discovery and Narrative Hypotheses. *Journal of Educational Psychology* 104(1): 235–249.

Arroyo, I.; Cooper, D.; Burleson, W.; Woolf, B.; Muldner, K.; and Christopherson, R. 2009. Emotion Sensors Go to School. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence in Education*, 17–24. Amsterdam: IOS Press.

Azevedo, R.; Moos, D. C.; Johnson, A. M.; and Chauncey, A. D. 2010. Measuring Cognitive and Metacognitive Regulatory Processes During Hypermedia Learning: Issues and Challenges. *Educational Psychologist* 45(4): 210–223.

Baker, R. S. J. D.; Goldstein, A. B.; and Heffernan, N. T. 2011. Detecting Learning Moment-By-Moment. *International Journal of Artificial Intelligence in Education* 21(1–2): 5–25.

Biswas, G.; Jeong, H.; Kinnebrew, J. S.; Sulcer, B.; and Roscoe, R. 2010. Measuring Self-Regulated Learning Skills Through Social Interactions in a Teachable Agent Environment. *Research and Practice in Technology Enhanced Learning* 5(2): 123–152.

Blaylock, N., and Allen, J. 2003. Corpus-Based, Statistical Goal Recognition. in *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence*, 1303–1308. San Francisco: Morgan Kaufmann Publishers.

Charniak, E., and Goldman, R. P. 1993. A Bayesian Model of Plan Recognition. *Artificial Intelligence* 64(1): 53–79.

Conati, C., and MacLaren, H. 2009. Empirically Building and Evaluating a Probabilistic Model of User Affect. *User Modeling and User-Adapted Interaction* 19(3): 267–303.

Craig, S.; Graesser, A.; Sullins, J.; and Gholson, B. 2004. Affect and Learning: An Exploratory Look into the Role of Affect in Learning with Autotutor. *Learning, Media and Technology* 29(3): 241–250.

Dean, T., and Kanazawa, K. 1989. A Model for Reasoning About Persistence and Causation. *Computational Intelligence* 5(3): 142–150.

de Vicente, A., and Pain, H. 2002. Informing the Detection of the Students' Motivational State: An Empirical Study. In *Proceedings of the Sixth International Conference on Intelligent Tutoring Systems*, 933–943. Berlin: Springer.

D'Mello, S., and Graesser, A. 2010. Multimodal Semi-Automated Affect Detection from Conversational Cues, Gross Body Language, and Facial Features. *User Modeling and User-Adapted Interaction* 20(2): 147–187.

Druzdzel, M. J. 1999. Smile: Structural Modeling, Inference, and Learning Engine and Genie: A Development Environment for Graphical Decision-Theoretic Models. In *Proceedings of the Eleventh Conference on Innovative Applications of Artificial Intelligence*, 902–903. Menlo Park, CA: AAAI Press.

Elliot, A. J., and Pekrun, R. 2007. Emotion in the Hierarchical Model of Approach-Avoidance Achievement Motivation. In *Emotion in Education*, ed. P. A. Schutz, and R. Pekrun, 57–74. Amsterdam: Elsevier.

Feng, M.; Heffernan, N.; and Koedinger, K. 2009. Addressing the Assessment Challenge with an Online System that Tutors as It Assesses. *User Modeling and User-Adapted Interaction* 19(3): 243–266.

Forbes-Riley, K., and Litman, D. 2011. Designing and Evaluating a Wizarded Uncertainty-Adaptive Spoken Dialogue Tutoring System. *Computer Speech and Language* 25(1): 105–126.

Gee, J. P. 2007. *What Video Games Have to Teach Us About Learning and Literacy*, 2nd ed. New York: Palgrave Macmillan.

Ha, E. Y.; Rowe, J. P.; Mott, B. W.; and Lester, J. C. 2011. Goal Recognition with Markov Logic Networks for Player-Adaptive Games. In *Proceedings of the Seventh International Conference on Artificial Intelligence and Interactive Digital Entertainment*, 32–39. Menlo Park, CA: AAAI Press.

Habgood, M. P. J., and Ainsworth, S. E. 2011. Motivating Children to Learn Effectively: Exploring the Value of Intrinsic Integration in Educational Games. *Journal of the Learning Sciences* 20(2): 169–206.

Hu, D. H., and Yang, Q. 2008. Cigar: Concurrent and Interleaving Goal and Activity Recognition. In *Proceedings of the Twenty-Third National Conference on Artificial Intelligence*, 1363–1368. Menlo Park, CA: AAAI Press.

Johnson, W. L. 2010. Serious Use of a Serious Game for Language Learning. *International Journal of Artificial Intelligence in Education* 20(2): 175–195.

Kautz, H., and Allen, J. F. 1986. Generalized Plan Recognition. In *Proceedings of the Fifth National Conference on Artificial Intelligence*, 32–37. Menlo Park, CA: AAAI Press.

Ketelhut, D. J.; Dede, C.; Clarke, J.; and Nelson, B. C. 2010. A Multi-User Virtual Environment for Building Higher Order Inquiry Skills in Science. *British Journal of Educational Technology* 41(1): 56–68.

Kim, J.; Hill, R. W.; Durlach, P. J.; Lane, H. C.; Forbell, E.; Core, M.; Marsella, S. C., Et Al. 2009. Bilat: A Game-Based Environment for Practicing Negotiation in a Cultural Context. *International Journal of Artificial Intelligence in Education* 19(3): 289–308.

Kirschner, P.; Sweller, J.; and Clark, R. 2006. Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist* 41(2): 75–86.

Koedinger, K. R.; Anderson, J. R.; Hadley, W. H.; and Mark, M. A. 1997. Intelligent Tutoring Goes to School in the Big City. *International Journal of Artificial Intelligence in Education* 8(1): 30–43.

Lee, S. Y.; Mott, B. W.; and Lester, J. C. 2012. Real-Time Narrative-Centered Tutorial Planning for Story-Based Learning. In *Proceedings of the Eleventh International Conference on Intelligent Tutoring Systems*, 476–481. Berlin: Springer.

Lee, S. Y.; Mott, B. W.; and Lester, J. C. 2011. Modeling Narrative-Centered Tutorial Decision Making in Guided Discovery Learning. In *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Education*, 163–170. Berlin: Springer.

Lim, M. Y.; Dias, J.; Aylett, R.; and Paiva, A. 2012. Creating Adaptive Affective Autonomous NPCS. *Autonomous Agents and Multi-Agent Systems* 24(2): 287–311.

Lin, G. I., and Walker, M. A. 2011. All the World's a Stage: Learning Character Models from Film. In *Proceedings of the Seventh Artificial Intelligence and Interactive Digital Entertainment Conference*, 46–52. Palo Alto, CA: AAAI Press.

Marsella, S. C., and Gratch, J. 2009. EMA: A Process Model of Appraisal Dynamics. *Cognitive Systems Research* 10(1): 70–90.

Mateas, M., and Stern, A. 2005. Structuring Content in the Facade Interactive Drama Architecture. In *Proceedings of Artificial Intelligence and Interactive Digital Entertainment Conference*, 93–98. Palo Alto CA: AAAI Press.

Mayer, R. 2004. Should There Be a Three-Strike Rule Against Pure Discovery Learning? *American Psychologist* 59(1): 4–19.

McCoy, J.; Treanor, M.; Samuel, B.; Wardrip-Fruin, N.; and Mateas, M. 2011. Comme Il Faut: A System for Authoring Playable Social Models. In *Proceedings of the Seventh International Conference on Artificial Intelligence and Interactive Digital Entertainment*, 158–163. Palo Alto, CA: AAAI Press.

Mcquiggan, S.; Lee, S.; and Lester, J. 2007. Early Prediction of Student Frustration. In *Proceedings of the Second International Conference on Affective Computing and Intelligent Interaction*, 698–709. Berlin: Springer.

Meluso, A.; Zheng, M.; Spires, H. A.; and Lester, J. 2012. Enhancing Fifth Graders' Science Content Knowledge and Self-Efficacy Through Game-Based Learning. *Computers and Education* 59(2): 497–504.

Mott, B. W., and Lester, J. C. 2006. U-Director: A Decision-Theoretic Narrative Planning Architecture for Storytelling Environments. In *Proceedings of the 5th International Joint Conference on Autonomous Agents and Multiagent Systems*, 997–984. New York: Association for Computing Machinery.

Picard, R. W.; Papert, S.; Bender, W.; Blumberg, B.; Breazeal, C.; Cavallo, D.; Machover, T. 2004. Affective Learning — A Manifesto. *BT Technology Journal* 22(4): 253–269.

Porteous, J.; Teutenberg, J.; Charles, F.; and Cavazza, M. 2011. Controlling Narrative Time in Interactive Storytelling. In *Proceedings of the Tenth International Conference on Autonomous Agents and Multiagent Systems*, 449–456. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems.

Richardson, M., and Domingos, P. 2006. Markov Logic Networks. *Journal of Machine Learning* 62(1–2): 107–136.

Riedel, S. 2008. Improving the Accuracy and Efficiency of Map Inference for Markov Logic. In *Proceedings of the Twenty-Fourth Annual Conference on Uncertainty in Artificial Intelligence*, 468–475. Seattle, WA: AUAI Press.

Riedl, M. O.; Stern, A.; Dini, D. M.; and Alderman, J. M. 2008. Dynamic Experience Management in Virtual Worlds for Entertainment, Education, and Training. *International Transactions on Systems Science and Applications, Special Issue on Agent Based Systems for Human Learning* 4(2): 23–42.

Riedl, M. O., and Young, R. M. 2004. An Intent-Driven Planner for Multi-Agent Story Generation. In *Proceedings of the Third International Conference on Autonomous Agents and Multi Agent Systems*, 186–193. Piscataway, NJ: Institute of Electrical and Electronics Engineers.

Rowe, J. P., and Lester, J. C. 2010. Modeling User Knowledge with Dynamic Bayesian Networks in Interactive Narrative Environments. In *Proceedings of the Sixth Annual Artificial Intelligence and Interactive Digital Entertainment Conference*, 57–62. Palo Alto, CA: AAAI Press.

Rowe, J. P.; Shores, L. R.; Mott, B. W.; and Lester, J. C. 2011. Integrating Learning, Problem Solving, and Engagement in Narrative-Centered Learning Environments. *International Journal of Artificial Intelligence in Education* 21(2): 115–133.

Sabourin, J. L.; Mott, B. W.; and Lester, J. C. 2011. Modeling Learner Affect with Theoretically Grounded Dynamic Bayesian Networks. In *Proceedings of the Fourth International Conference on Affective Computing and Intelligent Interaction*, 286–295. Berlin: Springer.

Sadilek, A., and Kautz, H. 2010. Recognizing Multi-Agent Activities from Gps Data. In *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence*, 1134–1139. Menlo Park, CA: AAAI Press.

Si, M.; Marsella, S.; and Pynadath, D. 2009. Directorial Control in a Decision-Theoretic Framework for Interactive Narrative. In *Proceedings of the Second International Conference on Interactive Digital Storytelling*, 221–233. Berlin: Springer.

Singla, P., and Mooney, R. 2011. Abductive Markov Logic for Plan Recognition. In *Proceedings of the Twenty-Fifth Conference on Artificial Intelligence*, 1069-1075. San Francisco, CA: AAAI Press.

Swanson, R., and Gordon, A. 2008. Say Anything: A Massively Collaborative Open Domain Story Writing Companion. In *Proceedings of the 1st Joint International Conference on Interactive Digital Storytelling*, 32–40. Berlin: Springer.

VanLehn, K. 2006. The Behavior of Tutoring Systems. *International Journal of Artificial Intelligence in Education* 16(3): 227–265.

Warren, S. J.; Dondlinger, M. J.; and Barab, S. A. 2008. A MUVE Towards Pbl Writing: Effects of a Digital Learning Environment Designed to Improve Elementary Student Writing. *Journal of Research on Technology in Education* 41(1): 113–140.

Woolf, B. P. 2008. *Building Intelligent Interactive Tutors: Student-Centered Strategies for Revolutionizing E-Learning*. San Francisco: Morgan Kaufmann Publishers.

Yu, H., and Riedl, M. O. 2012. A Sequential Recommendation Approach for Interactive Personalized Story Generation. In *Proceedings of the Eleventh International Conference on Autonomous Agents and Multiagent Systems*, 71–78. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems.

James Lester is a distinguished professor of computer science at North Carolina State University. His research focuses on transforming education with technology-rich learning environments. Utilizing AI, game technologies, and computational linguistics, he designs, develops, fields, and evaluates next-generation learning technologies for K–12 science, literacy, and computer science education. His work on personalized learning ranges from game-based learning environments and intelligent tutoring systems to affective computing, computational models of narrative, and natural language tutorial dialogue. He has served as program chair for the International Conference on Intelligent Tutoring Systems, the International Conference on Intelligent User Interfaces, and the International Conference on Foundations of Digital Games, and as editor-in-chief of the *International Journal of Artificial Intelligence in Education*. He has been recognized with a National Science Foundation CAREER Award and several Best Paper Awards.

Eun Ha is a postdoctoral research scholar in the Department of Computer Science at North Carolina State University. Her research interests lie in bringing robust solutions to complex problems at the interface of humans and computers by leveraging and advancing AI to create intelligent interactive systems. Her research in natural language has focused on discourse processing and natural language dialogue systems, and her research on user modeling has focused on goal recognition.

Seung Lee received the Ph.D. degree in computer science from North Carolina State University in 2012. His research interests are in multimodal interaction, interactive narrative, statistical natural language processing, and affective computing. He is currently with the SAS Institute where his work focuses on sentiment analysis in natural language processing.

Bradford Mott is a senior research scientist at the Center for Educational Informatics in the College of Engineering at North Carolina State University. His research interests include AI and human-computer interaction, with applications in educational technology. In particular, his research focuses on game-based learning environments, intelligent tutoring systems, computer games, and computational models of interactive narrative. Prior to joining North Carolina State University, he worked in the games industry developing cross-platform middleware solutions for the PlayStation 3, Wii, and Xbox 360. He received his Ph.D. from North Carolina State University in 2006.

Jonathan Rowe is a senior research associate in the Department of Computer Science at North Carolina State University. His research investigates applications of artificial intelligence and data mining for advanced learning technologies, with an emphasis on game-based learning environments. He is particularly interested in modeling how learners solve problems and experience interactive narratives in virtual worlds.

Jennifer Sabourin is a doctoral student in the Department of Computer Science at North Carolina State University. Her research focus is on developing effective educational technology with the application of AI and data mining techniques. She is particularly interested in the affective and metacognitive processes associated with learning with technology.