

# A Supervised Learning Framework for Modeling Director Agent Strategies in Educational Interactive Narrative

Seung Y. Lee, Jonathan P. Rowe, Bradford W. Mott, James C. Lester

**Abstract**—Computational models of interactive narrative offer significant potential for creating educational game experiences that are procedurally tailored to individual players and support learning. A key challenge posed by interactive narrative is devising effective director agent models that dynamically sequence story events according to players’ actions and needs. In this paper, we describe a supervised machine learning framework to model director agent strategies in an educational interactive narrative, CRYSTAL ISLAND. Findings from two studies with human participants are reported. The first study utilized a Wizard-of-Oz paradigm where human “wizards” directed participants through CRYSTAL ISLAND’s mystery storyline by dynamically controlling narrative events in the game environment. Interaction logs yielded training data for machine learning the conditional probabilities of a dynamic Bayesian network (DBN) model of the human wizards’ directorial actions. Results indicate that the DBN model achieved significantly higher precision and recall than naïve Bayes and bigram model techniques. In the second study, the DBN director agent model was incorporated into the run-time version of CRYSTAL ISLAND, and its impact on students’ narrative-centered learning experiences was investigated. Results indicate that machine learning director agent strategies from human demonstrations yields models that positively shape players’ narrative-centered learning and problem-solving experiences.

**Index Terms**—Narrative, interactive drama, serious games, Bayesian networks, machine learning.

## I. INTRODUCTION

RECENT years have witnessed substantial growth in research on computational models of interactive narrative in digital games [1–5]. Computational models of interactive

narrative aim to procedurally adapt story experiences in response to players’ actions, as well as tailor story elements to individual players’ preferences and needs. A common metaphor for interactive narrative models is a *director agent* (drama manager), which is a centralized software agent that works behind the scenes to procedurally direct a cast of non-player characters and storyworld events [4], [6–7]. The capacity to augment and revise narrative plans at run-time has shown promise for several applications, including entertainment [8–10], art [1], training [11], and education [6], [12–13]. In education, computational models of interactive narrative have been embedded in narrative-centered learning environments for a range of subjects, including language and culture learning [14], social skills development [12], network security [15], and middle school science [16].

Modeling interactive narrative director agents poses several computational challenges. Interactive narratives correspond to spaces of possible story experiences that grow exponentially in their number of events. Consequently, interactive narrative director agents are expected to effectively navigate large story spaces during run-time, and they are expected to recognize and react to players’ subjective experiences. In educational applications of interactive narrative, personalizing story events to support student learning and engagement is a promising direction for improving educational outcomes, but there is a dearth of theoretical guidance, or even best practices, to guide the design of interactive narrative models. In recognition of these challenges, efforts have been undertaken to automatically induce computational models of interactive narrative from large data sets [7], [17–18].

A promising approach for automating the creation of director agents is machine-learning models directly from human demonstrations. This approach requires humans to simulate director agents by controlling story events in an interactive narrative. Non-director players, often adopting the role of the story’s protagonist, simultaneously explore the narrative environment under the guidance of the human director’s actions. Data from these human-human interactions yields a training corpus for automatically inducing models of director agent strategies, which can be obtained by applying supervised machine learning techniques. The end result of this approach is a data-driven director agent model that can replace the human director in the interactive narrative environment. In educational interactive narratives, the director agent acts as a

Manuscript received June 14, 2013. This work was supported in part by the National Science Foundation under Grants REC-0632450 and DRL-0822200. 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. Additional support was provided by the Bill and Melinda Gates Foundation, the William and Flora Hewlett Foundation, and EDUCAUSE.

S. Y. Lee is with the Computer Science Department, North Carolina State University, Raleigh, NC 27695-8206 USA. (e-mail: [sylee@ncsu.edu](mailto:sylee@ncsu.edu)).

J. P. Rowe is with the Computer Science Department, North Carolina State University, Raleigh, NC 27695-8206 USA. (e-mail: [jprowe@ncsu.edu](mailto:jprowe@ncsu.edu)).

B. W. Mott is with the Computer Science Department, North Carolina State University, Raleigh, NC 27695-8206 USA. (e-mail: [bwmott@ncsu.edu](mailto:bwmott@ncsu.edu)).

J. C. Lester is with the Computer Science Department, North Carolina State University, Raleigh, NC 27695-8206 USA. (e-mail: [lester@ncsu.edu](mailto:lester@ncsu.edu)).

*narrative-centered tutor*, providing personalized guidance and feedback within the narrative to enhance the learning environment’s pedagogical effectiveness.

In this paper, we present a framework for machine learning director agent strategies from observations of human-human interactions in an educational interactive narrative. The approach involves training dynamic Bayesian network models of director agent strategies from a corpus of human directorial actions. In order to investigate the framework, we use a testbed interactive narrative called CRYSTAL ISLAND. CRYSTAL ISLAND features a science mystery where players investigate a spreading illness afflicting a team of scientists. Results from two empirical studies with human participants playing CRYSTAL ISLAND are reported. The first study employed a Wizard-of-Oz paradigm—in other words it involved *regular* (i.e., *non-wizard*) users interacting with *wizard* users providing directorial guidance in CRYSTAL ISLAND—in order to generate a training corpus for inducing director agent models. The corpus was used to machine learn conditional probability tables in a dynamic Bayesian network (DBN) model of the human wizards’ directorial strategies. The second study involved a modified version of CRYSTAL ISLAND that replaced human wizards with the DBN director agent model. A comparison between the DBN model and a baseline system is described, including the systems’ differential impacts on players’ narrative-centered learning experiences. Empirical findings demonstrating our framework’s impact on students’ learning outcomes are presented.

## II. RELATED WORK

Several families of algorithms have been employed for modeling interactive narrative director agents. Classical planning is one prevalent approach because STRIPS-style plans align naturally with computational representations of stories. Plan-based director agents monitor and revise the executions of story plans in order to respond to players’ actions and preserve desirable narrative properties [8]. In addition to classical planning, reactive planners have been investigated for dynamically responding to player actions under real-time performance constraints. Several of these systems incorporate special-purpose data structures inspired by narrative concepts, such as dilemmas or beats, in order to bundle story content for reactive delivery [1], [19].

Search-based approaches have been investigated for dynamically managing interactive narratives. Search-based approaches attempt to find plot sequences that optimize designer-specified evaluation functions [20]. These formalisms often use memoization or depth-bounded search techniques in order to constrain their computation times. However, they are sensitive to the narrative evaluation functions employed, which may be difficult to craft.

Case-based reasoning techniques have been used in several story-centric interactive narrative systems. The OPIATE story director dynamically responds to user actions by retrieving Proppian sub-plots rooted in particular story contexts using k-nearest neighbor techniques [21]. Work by Sharma et al. [9] modifies earlier search-based drama management approaches [20] by incorporating a case-based player model that

approximates users’ plot preferences.

Another important class of narrative adaptation techniques relies on decision-theoretic planning algorithms [3], [6–7], [22]. A family of interactive narrative models, known as declarative optimization-based drama managers (DODM), employs Markov decision processes to encode director agent tasks [7], [22]. DODM models’ parameters are automatically induced using on-line reinforcement learning techniques (such as temporal-difference learning) with large interactive narrative corpora generated from simulated users.

U-Director is an example of a decision-theoretic director agent that utilized dynamic decision networks in an early version of the CRYSTAL ISLAND interactive narrative [6]. The director agent triggers story-based hints that assist students while they investigate the interactive science mystery. THESPIAN is another example of a decision-theoretic interactive narrative system; it endows virtual characters with goal-oriented decision-making models that are loosely based on partially observable Markov decision processes (POMDPs) [3]. Each virtual character implements a recursive “theory of mind” model in order to reason about how its actions impact the beliefs and goals of other agents.

Recently, data-driven techniques have been employed to automatically devise interactive narrative models. Interactive narrative models have been machine learned from simulation datasets [7], [22], induced from large corpora of narrative blog entries [18], and distilled from crowd-sourced narrative descriptions [17]. In work that is perhaps most closely related to our own, Orkin [23] automatically induces models of social behavior and dialogue from thousands of human-human interactions in The Restaurant Game. Orkin’s approach, called collective artificial intelligence, differs from ours by focusing on models of characters’ social behaviors rather than director agent strategies. Further, Orkin’s computational framework combines crowd-sourcing, pattern discovery, and case-based planning techniques, whereas our approach leverages Wizard-of-Oz studies and dynamic Bayesian networks to induce models of director agent strategies.

Work on computational models of interactive narrative for education has also examined several algorithmic techniques. Efforts to provide personalized, story-embedded support for learning have largely focused on rule-based techniques for delivering hints [13], compositional partial-order planning techniques for scaffolding problem solving [15], hand-authored dynamic decision networks for narrative-centered tutorial planning [6], and POMDP-based multi-agent simulations for interactive pedagogical dramas [14]. Other work has examined character-based models of interactive narrative generation for social skills learning [12], although this work did not investigate director agents. In our work, we induce educational interactive narrative models directly from data generated by human users serving as director agents.

## III. DATA-DRIVEN DIRECTOR AGENT FRAMEWORK

Data-driven frameworks for learning models of director agent strategies hold considerable appeal for addressing the challenges inherent in creating director agent models. Our framework for inducing director agent models consists of four

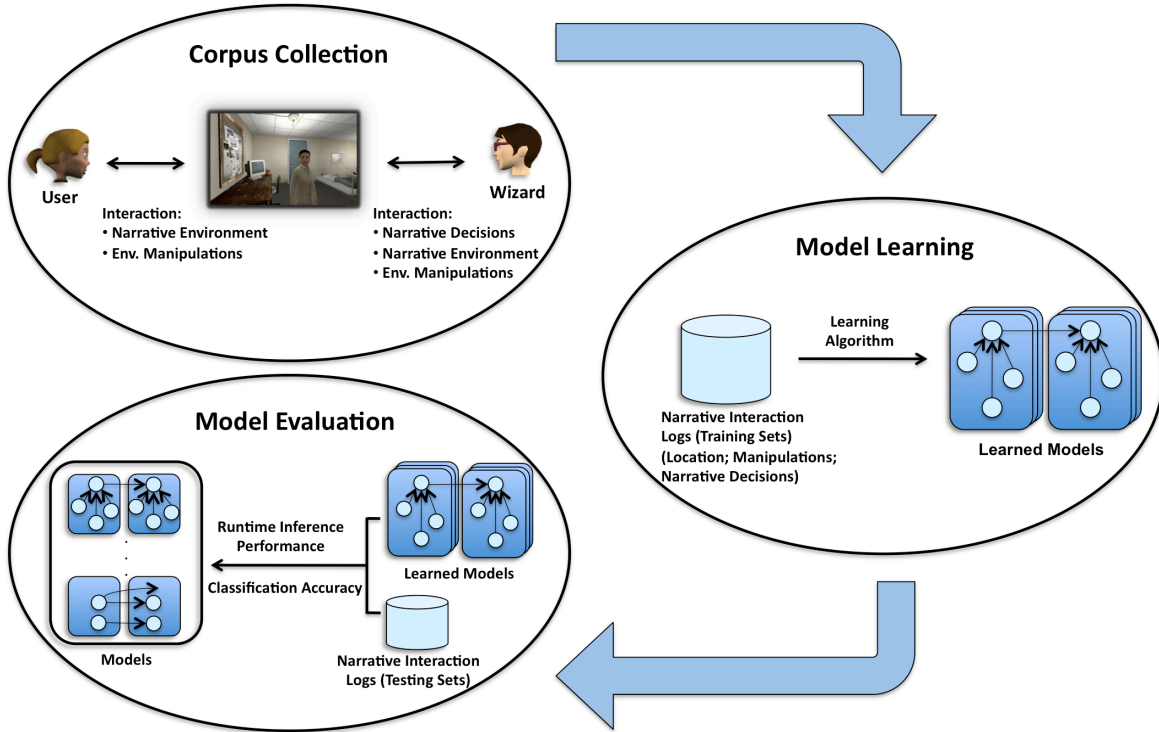


Fig. 1. Data-driven framework for modeling director agent strategies from human demonstrations.

phases: corpus collection, model learning, predictive model evaluation, and runtime model evaluation. Fig. 1 depicts the first three stages of the approach for inducing director agent strategies from human-human interaction data using dynamic Bayesian networks.

#### A. CORPUS COLLECTION

In order to acquire a corpus of interactive narrative training data that effectively represents canonical directorial sequences, the corpus collection process should address the following considerations. First, the corpus collection should be conducted using a data-collection version of the interactive narrative environment in which a human player, or wizard, controls the behavior of a simulated director agent. Other than the decision-making mechanism responsible for driving the director agent, the narrative environment should mirror the final “director-enhanced” version of the game; it is important for players’ interactions to closely resemble the interactions that ultimately occur in the final narrative environment. This symmetry increases the chances that a model trained from corpus data will also perform well when integrated into the final interactive narrative.

Second, the human director agent should be provided with an easy-to-use *narrative dashboard* in order to control the progression of the interactive narrative. The narrative dashboard enables the human director agent to perform actions (e.g., directing a non-player character to perform a specific action, triggering an in-game object to appear) in an omnipotent manner, mimicking the capabilities of a computational director agent. In addition to controls, the narrative dashboard should report on the storyworld’s state in

order to inform the wizard’s decisions. It is important to note that the dashboard must be designed carefully so as not to be too complex for the wizard to effectively use while an interactive narrative is underway.

Third, the wizard should make her directorial decisions based on observable story information brokered by the game environment (e.g., storyworld state, player behaviors), constraints imposed on the interactive story by the plot graph, and beliefs about actions that will lead to the most compelling story for the player.

Fourth, wizards should be given ample opportunity to familiarize themselves with the narrative-dashboard and interactive narrative’s structure as encoded by the plot graph. Pilot data collections can be performed as part of training wizards prior to collecting actual corpus data. One should not necessarily expect that a human will be an effective director agent without prior training or opportunities to familiarize herself with the narrative environment.

#### B. MODEL LEARNING

Models of director agent strategies are machine learned from narrative interaction logs acquired during a corpus collection study with human participants. Models are induced from training datasets using supervised machine learning techniques specific to the models being devised (e.g., dynamic Bayesian networks, naïve Bayes). The wizard’s directorial actions serve as class labels to be predicted by induced models. In choosing models to encode director agent strategies, several factors should be considered. First, the model should be capable of explicitly representing changes in the director agent’s belief state over time, as well as temporal



Fig. 2. WOZ-enabled CRYSTAL ISLAND.

changes to the game’s narrative and world states. Second, the director agent model should not only recommend what directorial action to perform at a given decision point, but also indicate the appropriate time to intervene. Third, the director agent should be capable of integrating observations from several sources to inform its directorial strategies. Sources of observations might include narrative history, storyworld state, player activity, and player beliefs. Finally, the model must be capable of addressing these requirements while operating at run-time, functioning efficiently enough for integration with game engine technologies.

### C. PREDICTIVE MODEL EVALUATION

Once a director agent model is induced using a training corpus and supervised machine learning algorithms, the learned model should be evaluated using test datasets to examine the director agent model’s ability to predict a human director’s narrative decisions. The performance of the learned model can be evaluated with respect to predictive accuracy, including metrics such as precision and recall. The model should be compared to a baseline approach in order to determine how effectively the learned model performs relative to alternate techniques.

### D. RUNTIME MODEL EVALUATION

Once a sufficient level of predictive accuracy is obtained, the director agent model can be integrated into the runtime interactive narrative system. This introduces the opportunity to empirically evaluate the induced director agent model through studies with human participants examining how the system affects and engages players.

## IV. CRYSTAL ISLAND INTERACTIVE NARRATIVE TESTBED

To investigate director agent strategies, a Wizard-of-Oz<sup>1</sup> data collection was conducted with a customized version of the CRYSTAL ISLAND interactive narrative [16].

### A. CRYSTAL ISLAND

CRYSTAL ISLAND is an educational adventure game that features an interactive science mystery set on a recently discovered tropical island. The game is built using Valve

Corporation’s Source™ engine, the game technology behind the popular Half-Life® 2 series. CRYSTAL ISLAND has been designed to help middle school students learn microbiology concepts through an engaging and immersive story-centric experience [16]. Within the story, the player adopts the role of a protagonist attempting to discover the identity and source of an infectious disease plaguing a research station that is located on the island. Throughout the mystery, the player is free to explore the world and interact with other characters while forming questions, generating hypotheses, collecting data, and testing hypotheses. The player can pick up and store objects, view posters, operate lab equipment, and talk with non-player characters to gather clues about the source of the disease. During the course of solving the mystery, the player completes an in-game diagnosis worksheet to organize her findings, hypotheses, and conclusions. Upon completing the diagnosis worksheet, the player verifies its contents with the camp nurse and develops a treatment plan for the sickened CRYSTAL ISLAND team members.

### B. CRYSTAL ISLAND: Wizard-of-Oz Version

For the corpus collection, a custom episode of CRYSTAL ISLAND (Fig. 2) was created that includes a companion agent who assists the player to solve the mystery. The player adopts the role of Alex Reid visiting her father, Bryce, who serves as the research station’s lead scientist. CRYSTAL ISLAND’s narrative backstory is as follows: Alex has arrived at CRYSTAL ISLAND to visit her father whom she has not seen for a while. As she approaches the dock, she hears news that her father has fallen ill from AI, the camp foreman. AI tells her that Audrey, Ford, and her father were out on an expedition gathering specimens. Their expedition was scheduled to last for two days; however, they failed to return to the camp on time. AI found this very unusual since they were known to adhere closely to schedule. Fearful for their safety, AI led a search team to locate them. After two days of searching, the research team discovered that the expedition team had fallen ill on the south side of the island. It appears that the group lost their way, became ill, and could not make it back to the camp. They are in the infirmary and are being attended to by the camp’s nurse. Upon hearing the news, Alex hurries to the infirmary to see her father and his colleagues. Kim, the camp’s nurse, informs her that their condition is poor. Her father seems to be doing much worse than the others. Kim is baffled by the

<sup>1</sup> A Wizard-of-Oz study is a study paradigm in which participants interact with what appears to be an autonomous computer system, but is actually a person (the “wizard”) in another location.

illness and does not know what could have caused it. She asks Alex to help her identify the disease and its source.

### C. Wizard-of-Oz Functionalities

To investigate director agent strategies, CRYSTAL ISLAND was extended to include new functionalities specific to a Wizard-of-Oz study design. In this WOZ-enabled version of CRYSTAL ISLAND, a human wizard provides narrative planning functionalities as well as spoken natural language dialogue for the companion character. Playing the role of the camp nurse, the wizard works collaboratively with the player to solve the science mystery while also performing directorial actions to guide the interactive narrative experience. Together in the virtual environment, the wizard and player carry on conversations using voice chat and observe one another's actions while investigating the mystery scenario. In addition to directing the nurse character's navigation, spoken communication, and manipulation behaviors, the wizard guides the player's investigative activities and controls the narrative's pace and progression. To support these activities, the wizard's computer terminal includes a detailed dashboard that provides information about the player's activities in the environment (e.g., reading books, testing objects, updating the diagnosis worksheet) as well as controls to initiate key narrative events in the environment (e.g., introducing new patient symptoms, having a non-player character deliver additional items for testing). This narrative dashboard provides the wizard with sufficient capabilities to simulate a computational director agent.

In addition to new wizard functionalities, the CRYSTAL ISLAND narrative environment was streamlined to increase focus on a reduced set of narrative interactions between the player and wizard, as well as reduce the amount of time spent navigating the environment. This was accomplished by confining the scenario to a single building that encompasses the camp's infirmary and laboratory. Within this environment the player and wizard gain access to all of the materials needed to solve the science mystery (e.g., sickened researchers, background books and posters, potential sources of the disease, lab equipment). The scenario, controls, and narrative dashboard were refined throughout a series of pilot studies with college students that were conducted prior to the corpus collection described in this paper.

### D. Example Scenario

To illustrate the behavior of the WOZ-enabled CRYSTAL ISLAND environment, consider the following scenario. A player has been collaborating with the nurse character, whose behaviors are planned and executed by a human wizard. The player has learned that an infectious disease is an illness that can be transmitted from one organism to another, often through food or water. Under guidance of the nurse, the player has examined the patients' symptoms and run lab tests on food items. Through this exploration, the user has come to believe that the source of the illness is a waterborne disease and that it is likely cholera or shigellosis. Although she believes cholera is more likely, she is unable to arrive at a final diagnosis.

Through her conversation with the nurse character, the wizard determines that the player is having difficulty ruling out shigellosis. The wizard decides that this is an opportune moment to provide a hint. The wizard uses the narrative dashboard to enable the *Observe Leg Cramp Symptom* plot point, which results in one of the patients moaning loudly in the infirmary. The player hurriedly examines the patient and informs the wizard, "He has leg cramps. That means it is cholera." The wizard asks the player to update her diagnosis worksheet with her new hypothesis and explain why she believes her recent finding. The player then provides a detailed explanation justifying her diagnosis, and the story concludes with the nurse treating the patients for cholera.

## V. CORPUS COLLECTION PROCEDURE

A study with human participants was conducted in which more than twenty hours of trace data were collected using the WOZ-enabled version of the CRYSTAL ISLAND game environment. The trace data includes detailed logs of all player and wizard actions (e.g., navigation, manipulation, and decision making) in the interactive narrative environment, as well as audio and video recordings of their conversations.

### A. Participants

The participants were 33 eighth-grade students (15 males and 18 females) from North Carolina ranging in age from 13 to 15 ( $M = 13.79$ ,  $SD = 0.65$ ). Two wizards assisted with the corpus collection, one male and one female. Each session involved a single wizard and a single student. The student and wizard were physically located in different rooms throughout the session.

### B. Wizard Protocol

To improve the consistency of the wizards' interactive narrative decision-making and natural language dialogue activities, a wizard protocol was iteratively developed and refined through a series of pilot studies. The resulting protocol included a high-level procedure for the wizards to follow (e.g., introduce yourself as the camp nurse, describe the patient situation to the player), a set of interaction guidelines (e.g., collaboratively work with the player to solve the mystery, encourage the player to explain her conclusions), and a set of narrative guidelines (e.g., descriptions about the overall story structure, suggestions about appropriate contexts for narrative decisions, explanations about event ordering constraints).

Prior to the corpus collection with the eighth grade students, each wizard was trained on the CRYSTAL ISLAND microbiology curriculum and the materials that would be provided to students during the corpus collection. The wizard training also included information on key concepts from the CRYSTAL ISLAND curriculum and the protocol to follow. After carefully reviewing the materials over the course of a week and having any of their questions answered, the wizards participated in at least three training sessions with college students. After each training session, a researcher performed an "after action review" with the wizard to discuss his or her interactions with the students and adherence to the wizard protocol. Through these training sessions, wizards gained



considerable experience performing directorial actions in CRYSTAL ISLAND, and they developed strategies for directing students toward successful learning and problem-solving outcomes. While the wizards were not professional tutors, they did have significant prior experience in educational technology. Consequently, the wizards were considered to possess above-novice pedagogical skills, as well familiarity with the CRYSTAL ISLAND environment. These qualifications suggested that the wizards were capable of demonstrating effective directorial strategies for machine learning.

### C. Participant Procedure

When participants arrived at the data collection room, they were greeted by a researcher and instructed to review a set of CRYSTAL ISLAND handouts, including information on the back story, task description, characters, and controls. Upon completing their review of the handouts, the researcher provided further direction to the participants on the use of the keyboard and mouse controls. The researcher then informed the participants that they would be collaborating with another human-controlled character, the camp nurse, in the environment to solve the science mystery. Participants were asked to communicate with the camp nurse throughout their sessions. Finally, the researcher answered any questions from the participants, informed them that the sessions were being videotaped, instructed them to put on their headsets and position their microphones, and asked them to direct all future communication to the camp nurse. The researcher remained in the room with the participant for the duration of their session. The CRYSTAL ISLAND session concluded once the participant and wizard arrived at a treatment plan for the sickened virtual scientists. The participants' sessions lasted no more than sixty minutes. During data analysis, data from one of the participants was eliminated as an outlier—the data was more than three standard deviations from the mean—leaving thirty-two usable trace data logs.

## VI. MODELING DIRECTOR AGENT STRATEGIES WITH DYNAMIC BAYESIAN NETWORKS

The Wizard-of-Oz study yielded a training corpus for machine learning models of director agent strategies in the CRYSTAL ISLAND narrative environment. In order to emulate director agent strategies effectively, it is necessary to induce models that prescribe when to *intervene* with a directorial action, as well as which *action* to perform during an intervention. These two tasks correspond to two sub-models for the director agent: an intervention model and an action model. Director agents utilize numerous storyworld observations that change over time to accurately determine the most appropriate time to intervene and the next director agent action to perform in an unfolding story.

In order to induce models to perform these tasks, we employed dynamic Bayesian networks [24]. A dynamic Bayesian network (DBN) is a directed acyclic graph that encodes a joint probability distribution over states of a complex system over time. DBNs consist of nodes representing random variables, directed links that encode

conditional dependencies among random variables, and conditional probability distributions annotating the nodes. A defining feature of DBNs is their use of *time slices*, which characterize the state of the underlying system at particular phases of time. By utilizing time slices, DBNs support probabilistic inference about events that change over time.

The DBN models in this work were implemented with the GeNIe/SMILE Bayesian modeling and inference library [25]. The DBNs' network structure was hand authored, but the conditional probability tables that annotate each node were automatically induced from the training corpus. The Expectation-Maximization algorithm from the SMILearn library was used to learn the DBNs' conditional probability tables. Expectation-Maximization was used as a matter of convenience; there were no hidden variables or missing data (i.e., missing attributes). SMILearn's implementation of Expectation-Maximization served as an off-the-shelf method for parameter learning. More specialized parameter-training techniques, such as relative-frequency estimation, would have also been appropriate, but a single iteration of Expectation-Maximization can serve the same purpose for fully observable models such as the DBNs in this work. After the models' parameters were learned, the resulting networks were used to infer directorial decisions about when to intervene in the CRYSTAL ISLAND interactive narrative, as well as which action to perform during an intervention. It should be noted that while the set of conditional probability values induced as parameters for the DBN models are specific to the story arc, characters, and game environment of CRYSTAL ISLAND, we anticipate that the data-driven methodology advocated by this paper generalizes across narrative spaces and game environments.

### A. Feature Selection

Feature selection is an important problem that has been studied for many years by the machine learning community [26]. By selecting the most relevant predictor features from a corpus of training data, the performance of machine learned models are often improved considerably. In our study we utilized a two-step feature selection approach to induce accurate models of director agents' intervention and action decision strategies. First, we performed a post-study "after action review" with the wizards to discuss his/her narrative decision strategies while interacting with the players. During the corpus collection we video recorded all of the actions taken by the wizards within the WOZ-enabled version of CRYSTAL ISLAND. After the corpus collection was over, we asked the wizards to watch the recorded videos and "think aloud" by retrospectively describing why they chose particular narrative decisions. The wizard was periodically prompted to explain the factors that drove her directorial decisions. The wizard was also asked what features she considered when making the narrative decisions. Afterward, we conducted a qualitative analysis of the wizards' responses to identify an candidate features for the intervention and action models.

During the second stage of feature selection, we performed a brute force feature ranking method. Selected features were

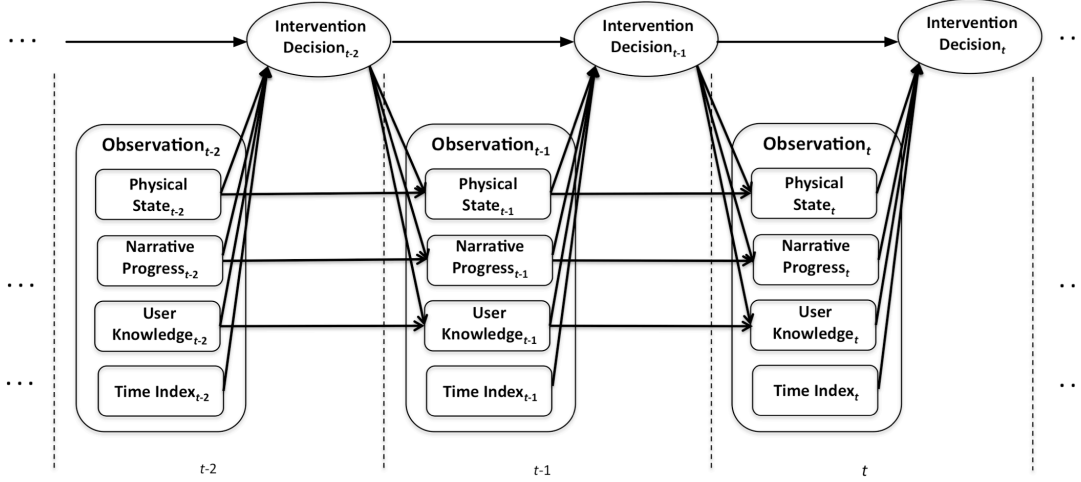


Fig. 3. High-level illustration of dynamic Bayesian network model for director agent intervention strategies.

evaluated with all possible combinations of the input features. Subsets consisting of features that yielded the most efficient performance for each intervention and action model were chosen and implemented. The initial pool consisted of eight features identified from the retrospective commentaries provided by wizards. The automated brute force feature selection process reduced the set to four features in the intervention model and two features in the action model. There were no features considered from the spoken natural language dialogues between wizards and players. For the purpose of this investigation, all features were computed from the in-game trace data.

### B. Intervention Strategies

The high-level structure of the dynamic Bayesian network model of director agent intervention strategies—i.e., when to perform a directorial action—is shown in Fig. 3. Three time-slices are illustrated in the figure. In each slice, the intervention decision from the previous time slice,  $intervention\ decision_{t-1}$ , influences the current intervention decision,  $intervention\ decision_t$ . Within each time slice, observations from the story world, collectively known as  $narrative\ state_t$ , also influence the intervention decision. These observations include the physical state of the storyworld, progression of the narrative, user knowledge of the story, and the overall story timeline. Each time slice encodes a probabilistic representation of the director agent’s beliefs about the overall state of the narrative.

The DBN director agent intervention model consists of the following variables:

- **Intervention Decision.** Intervention decision is a binary variable with two values: *action* and *no-action*. *Action* indicates that a director agent should perform an action to intervene in the story. *No-action* indicates that the director agent should remain inactive.
- **Physical State.** Physical state is a discrete variable, with nine possible values, that encodes the player’s current

location and the wizard’s location in the story world. Locations are subdivided into discrete regions spanning the environment’s virtual infirmary and laboratory. All user interactions occur within these locations.

- **Narrative Progress.** Narrative progress is a discrete variable, with five possible values, that characterizes the narrative structure of CRYSTAL ISLAND’s plot. To represent narrative progress, we modeled CRYSTAL ISLAND’s plot structure in terms of a five-phase narrative arc framework: *exposition*, *complication*, *escalation*, *climax*, and *resolution*. Transitions between narrative phases were deterministically triggered by the occurrence of plot points within CRYSTAL ISLAND’s narrative. For example, when the leg-cramp-reveal plot point occurred in the narrative, that marked the transition from the climax phase to the resolution phase, because at this point the student had all of the information needed to diagnose the illness with certainty. These triggers were manually specified by CRYSTAL ISLAND’s designer. Both students and wizards experienced the same real-time progression of narrative phases; there were no differences in how wizards and students witnessed the plot advance, although wizards were able to control how and when certain plot points occurred through directorial actions.
- **Player Knowledge.** Player knowledge nodes are discrete variables, with ten possible values, that encode the player’s beliefs about salient facts of the story learned through interactions with the narrative environment and non-player characters. Player knowledge is measured on an ordinal scale. Within the CRYSTAL ISLAND environment, players complete a *diagnosis worksheet* while solving the science mystery, which provides details regarding players’ current beliefs about the story. Player knowledge node values are determined from user performance on the *diagnosis worksheet*.
- **Time Index.** Time index nodes are discrete variables that model the overall timeline of the storyworld, providing

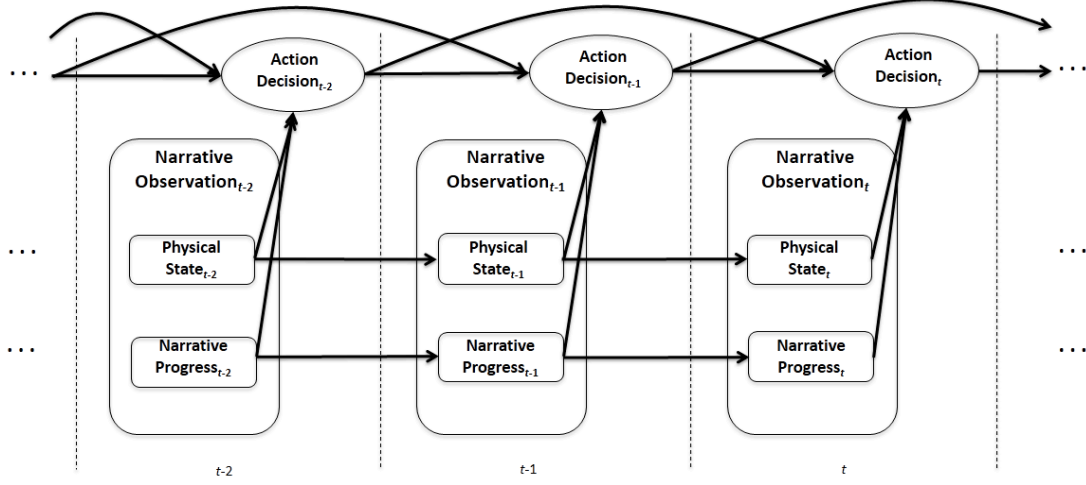


Fig. 4. High-level illustration of dynamic Bayesian network model for director agent action strategies.

temporal evidence to guide intervention decisions. There are 84 possible time index values, each corresponding to a discrete time slice in the model.

During runtime, new observations are made available to the director agent—such as player locations and narrative phases—which trigger corresponding nodes in the DBN to update with observed values. These updates propagate throughout the network by changing the marginal probabilities of connected nodes. Propagating evidence enables the model to perform inferences about the most probable *intervention decision*<sub>*t*</sub> at time slice *t*, which encodes whether a director agent should intervene or not intervene at a given point in time.

### C. Action Strategies

The high-level structure of the dynamic Bayesian network model created for director agent action strategies—i.e., which directorial action to perform—is shown in Fig. 4. The set of directorial actions were those controlled by the wizard’s narrative dashboard; navigation, object manipulation, conversation, and non-player character behaviors were not part of the model. The figure illustrates three time slices and their corresponding narrative action decisions: *action decision*<sub>*t-2*</sub>, *action decision*<sub>*t-1*</sub>, and *action decision*<sub>*t*</sub>. The three time slices incorporate narrative observations that encode information about the physical state of the storyworld and the narrative’s plot progression. The action model shares structural similarities with the intervention model, although different predictor features are used.

In the action model, *action decision* nodes encode current and prior actions performed by the director agent during interventions. The *Physical state* and *Narrative progress* nodes are identical to the equivalent nodes in the intervention strategies model described in the prior section. When selecting directorial actions, the model consults beliefs about the narrative environment’s current physical state and plot progress, as well as its prior history of *action decision*<sub>*t-1*</sub> and *action decision*<sub>*t-2*</sub> (Fig. 4).

Given a DBN structure such as the one described above, conditional probability tables (CPTs) for each observation node in the network can be induced from a training corpus of wizard-player interaction data. After the model’s network structure and conditional probability tables have been obtained, the model can be used to guide director agent strategies during runtime. When the model is used for runtime decision-making, observed evidence is provided to the DBN model, which causes updates to marginal probability values in the network that in turn affect the computed probabilities of intervention and action decisions at each time slice.

### D. Predictive Model Evaluation Results and Discussion

Models of director agent intervention and action strategies were machine learned from the training corpus. The predictive performance of these models was evaluated with respect to predictive accuracy, recall, and precision using cross-validation techniques. This subsection describes results from these two machine-learning analyses.

#### Directorial Intervention Model

The unrolled DBN intervention model contained a total of 84 time slices. We chose 84 time slices by considering the average length of a narrative episode in the training corpus (approximately 45 minutes), identifying the minimum time required to perform a run-time Bayesian update without inhibiting CRYSTAL ISLAND’s graphical performance (approximately 32 seconds), and then computing the maximum number of Bayesian updates that could occur in an average episode ( $45 * 60 / 32 \approx 84$ ). The unrolled network consisted of 84 time slices. It should be noted that the large number of conditional probabilities in each time slice raised potential data sparsity concerns for machine learning the DBN’s parameters. While GeNIe requires the unrolled network be provided as an input for parameter learning, data sparsity issues were ameliorated by GeNIe’s parameter learning process: the parameters for a given time slice in a



DBN are learned using data from all time slices in the training corpus. In our case, 33 traces multiplied by 22 observations per slice yielded approximately 726 training observations per DBN time slice. In effect, the unrolled DBN’s parameter learning process mirrors training a DBN consisting of just two time slices. It is possible to further reduce data sparsity concerns by performing smoothing. However, we did not perform an explicit smoothing step in our analysis. The DBN’s conditional probability tables were directly those computed by GeNIe/SMILE’s implementation of Expectation-Maximization with the training corpus.

Statistical analyses were conducted to assess the effectiveness of dynamic Bayesian networks for modeling director agent intervention strategies. To determine the relative effectiveness of the DBN model, an alternate naïve Bayes model was developed as a baseline for comparison purposes. The naïve Bayes model leveraged the same set of observation variables as the DBN model, but the observable variables were assumed to be conditionally independent of one another. Further, temporal relations between successive time slices were not included in the naïve Bayes model. By contrast, the DBN model included conditional dependence relations between successive time slices. These relations are shown as directed links in Fig. 3. Both of the models were learned using the same training corpus: trace data collected from thirty-two player interactions with the CRYSTAL ISLAND game environment during the aforementioned Wizard-of-Oz study. A leave-one-out cross validation method was employed. We employed leave-one-out cross validation in order to ensure that we had enough training data to induce accurate predictive models. In some cases, leave-one-out cross validation can be subject to overfitting—a potential tradeoff of this validation scheme—but in practice we find that it is sufficient for validating models that perform well in run-time settings.

Recall, precision, and accuracy were computed for each model. Table I displays classification results for the naïve Bayes and DBN models. The DBN model outperformed the baseline naïve Bayes model in all categories, achieving a more than 16% absolute improvement in accuracy over the baseline approach. Also, the DBN model achieved sizable absolute gains in recall and precision, 34% and 50% respectively, as compared to the baseline approach.

The evaluation indicates that inducing dynamic Bayesian network models from human demonstrations of director agent intervention strategies is an effective approach, and the independence assumptions inherent in naïve Bayes models may not be appropriate for interactive narrative environments.

*Directorial Action Model*

The unrolled DBN model of director agent action strategies contained a total of 22 time slices. We chose 22 time slices because that was the maximum number of interventions observed in the training episodes. Once again, the DBN used conditional probability table values induced from GeNIe/SMILE’s implementation of Expectation-Maximization run on the training corpus. Statistical analyses were performed to investigate the effectiveness of dynamic Bayesian networks for predicting human directors’ action

TABLE I  
CLASSIFICATION RESULTS OF DIRECTOR AGENT INTERVENTION MODELS

Narrative Intervention Model	Recall	Precision	Accuracy
Naïve Bayes	39.3 %	31.9 %	75.5 %
DBN	73.3 %	82.0 %	92.8 %

TABLE II  
CLASSIFICATION RESULTS OF DIRECTOR AGENT ACTION MODELS

Narrative Action Model	Recall*	Precision*	Accuracy
bi-gram	74.0 %	73.0 %	71.0 %
DBN	93.0 %	94.0 %	93.7 %

\*Macro-averaged Recall and Precision

decisions during interactive narrative interventions. To examine the relative effectiveness of the DBN model compared to a baseline approach, a bi-gram model was developed in which only the previous action decision was used to predict the next action decision. This network structure was an appropriate baseline for comparing against the DBN model because it represents the most basic form of temporal probabilistic model for guiding director agent strategies.

The models’ predictive performances were evaluated in terms of overall accuracy, macro-averaged recall, and macro-averaged precision. The action decision model solves a multi-class classification problem, which requires the use of macro-averaging techniques when using evaluation metrics such as recall and precision [27]. Macro-averaging is commonly used by the natural language processing community to evaluate linguistic classifiers. Macro-averaging is performed by taking the arithmetic means of all predicted classes. This method assumes that each class has equal weight. The resulting values are considered macro-averaged recall and macro-averaged precision, respectively.

Leave-one-out cross validation was employed to evaluate the DBN and bi-gram director agent models. Results from the evaluation are shown in Table II. The bi-gram model predicted human directorial actions with 71% accuracy, whereas the DBN model achieved an accuracy of 93.7%. These results correspond to a 23% absolute improvement by the DBN model over baseline approaches. Further, the DBN model achieved sizable gains in both recall and precision compared to the baseline approach. These results suggest that leveraging evidence about narrative structure, physical locations, and action decision history can substantially improve models of director agents’ action decisions.

## VII. EVALUATING DBN MODELS OF DIRECTOR AGENT STRATEGIES IN RUNTIME SYSTEMS

To examine the effectiveness of DBN models of director agent strategies in a runtime setting, the intervention model and action model were both integrated into the CRYSTAL ISLAND interactive narrative environment. After the integration was completed, the updated version of CRYSTAL ISLAND was the subject of a second study involving human

participants. In this study, the game environment was identical to the WOZ-enabled version of CRYSTAL ISLAND from the corpus collection with two exceptions: 1) the DBN director agent controlled narrative events and non-player character behaviors rather than a human wizard, and 2) the nurse character’s dialogue was delivered in the form of pre-recorded speech and text rather than natural language conversation between the wizard and participant. While the omission of natural language dialogue is a limitation, we anticipated that it would not impact the interactive narrative’s dynamics in such a manner that would inhibit the DBN director agent’s ability to impact players’ narrative-centered learning experiences.

The director agent’s actions were reified in the narrative environment by the camp nurse character. When the directorial action was chosen, the nurse automatically moved to an appropriate location in the story world and performed the directed action. For example, if the director agent model chose to advance the interactive narrative by helping the player gather patient symptoms, the director agent would direct the camp nurse to approach the player and suggest that the player examine the patient. The camp nurse would also lead the player toward the infirmary where patients lay in medical cots. Whenever the nurse was in an idle state (i.e., not implementing a directorial action), she followed the user around the environment using built-in path finding, only responding to user requests for information. The camp nurse’s dialogue was presented through simultaneous speech and text; pre-recorded speech was provided a human voice actor, and the text of her dialogue appeared at the bottom of the screen. Players interacted with the nurse and other non-player characters to receive environmental information (e.g., How does one operate the laboratory equipment? Where is the library?), and uncover clues about the science mystery (e.g., What is a waterborne disease? What are Bryce’s symptoms?). Players selected their questions using dialogue menus. All character dialogue, including the nurse’s dialogue, was selected deterministically, and it was fully specified in fixed dialogue trees. In the case of the nurse, the dialogue tree was designed to approximate, at a simple level, the conversations that occurred during the Wizard-of-Oz corpus collection study. However, the DBN director agent model did not impact characters’ dialogue behavior, except in cases where triggering dialogue was part of a directorial action.

#### A. Runtime Evaluation Experiment

An evaluation experiment was conducted to examine players’ narrative learning experiences while interacting with the updated CRYSTAL ISLAND environment with integrated DBN director agent models. The between-group study design included two experimental conditions: one condition featured the DBN models of director agent strategies to guide narrative events in CRYSTAL ISLAND. The second condition used a simplified model that did not leverage machine learned directorial strategies. Participants’ narrative experiences and learning outcomes were compared between conditions to examine the relative effectiveness of director agent models induced from human demonstrations.

#### Machine-Learned Model

In the condition with machine-learned director agent models, players investigated the CRYSTAL ISLAND mystery under the guidance of the DBN director agent model. The director agent actively monitored players as they interacted with the storyworld, determined when to intervene, and selected appropriate directorial actions to guide participants through the intended narrative. The director agent had control over director intervention decisions (i.e., deciding when to intervene) and director action decisions (i.e., selecting what intervention to perform). The machine-learned director agent model had access to fifteen potential directorial actions—actions that were originally controlled through the narrative dashboard during the Wizard-of-Oz study—which are listed in Table III. The machine-learned model did not directly control non-player characters’ behavior, or the nurse’s navigation and dialogue behavior, except in cases where the behaviors were part of a directorial action listed in Table III.

#### Base Model

In the base condition, players investigated the CRYSTAL ISLAND mystery under the guidance of a minimal director agent model that did not use machine learned directorial strategies. This director agent model controlled a subset of 5 directorial actions (Table III) that were required for solving the mystery (i.e., the player could not progress in the narrative without the director taking action). The director agent in this condition was a simple rule-based model, not a machine-learned model. Directorial decisions such as introducing new patient symptoms and objects, were triggered whenever specific pre-conditions were satisfied in the environment.

This particular baseline was chosen for several reasons. First, we considered comparing the machine-learned model to a human wizard. However, it was logistically infeasible to recruit, train, and observe additional wizards (with no prior experience) to participate in the study due to the planned number of participants. Furthermore, human wizards would introduce a confounding variable—natural language dialogue—to the experiment that would potentially impact the results. Therefore, we removed this study design from consideration. Second, we considered comparing the machine-learned model to a prior director agent model for the CRYSTAL ISLAND environment, U-Director (Mott & Lester, 2006). However, U-Director was designed for an older version of CRYSTAL ISLAND, and it would have required a complete re-implementation to work in the updated environment. Consequently, we elected to pursue a third option, a rule-based model that provides minimal pedagogical guidance, to examine the DBN model’s ability to effectively coach students as they solved the science mystery.

#### B. Study Method

A total of 123 participants played CRYSTAL ISLAND. Participants were middle school students ranging in age from 12 to 15 ( $M = 13.40$ ,  $SD = 0.53$ ). Six of the participants were eliminated due to hardware and software issues, and fifteen participants were eliminated due to incomplete data on pre-

and post-experiment questionnaires. Among the remaining participants, 45 were male and 57 were female.

Several days prior to playing CRYSTAL ISLAND, participants completed a short series of pre-experiment questionnaires. The questionnaires included a microbiology curriculum test to assess student understanding of science concepts that were important in CRYSTAL ISLAND's science mystery. The curriculum test consisted of 20 multiple-choice questions about microbiology concepts, including the scientific method, pathogens, disease transmission methods, and infectious diseases. During the experiment, participants were randomly assigned to a study condition, and they were given 45 minutes to complete CRYSTAL ISLAND's interactive narrative scenario. Immediately after solving the mystery, or 45 minutes of interaction, whichever came first, participants exited the CRYSTAL ISLAND game environment and completed a series of post-experiment questionnaires. By terminating interactive narrative episodes after 45 minutes, it was ensured that the director agent models would not exceed the supported number of time slices. The post-experiment questionnaires included the microbiology curriculum test, which was identical to the pre-test version. The post-experiment questionnaires took approximately 30 minutes to complete. In total, the study sessions lasted no more than 90 minutes.

### C. Results

Statistical analyses were conducted to assess the relative effectiveness of the director agent models between participants. The analyses focused on two experiential outcomes: participants' science learning gains and participants' efficiency at solving the science mystery. The first outcome relates to how successfully the director agent model promotes CRYSTAL ISLAND's primary objective: science learning. The second outcome relates to how effectively the director agent model guides participants toward a desired narrative resolution.

Participants achieved significant positive learning gains from playing CRYSTAL ISLAND. A matched pairs *t*-test comparing pre- and post-experiment curriculum test scores indicated that participants' learning gains were statistically significant,  $t(102) = 2.23$ ,  $p < .05$ . Examining learning outcomes within each condition revealed that participants in the *Machine-Learned Model* condition achieved significant learning gains, but participants in the *Base Model* condition did not achieve significant gains (Table IV). Furthermore, there was a significant difference between the conditions in terms of learning gains. An analysis of covariance (ANCOVA) comparing post-test scores between conditions while controlling for pre-test scores revealed that learning gains from the *Machine-Learned Model* were significantly greater than the *Base Model*,  $F(2, 99) = 38.64$ ,  $p < .001$ .

A second analysis of the director agent model focused on participants' efficiency at solving CRYSTAL ISLAND's mystery. The investigation examined two metrics for each condition: whether participants solved the mystery, and the duration of time taken by participants to solve it. Table V shown means and standard deviations for the metrics in each condition.

TABLE III  
DIRECTOR AGENT DECISIONS

Decisions	Descriptions
START-SESSION*	Agent gives a brief explanation of the user's objectives and goals.
INTRODUCE-SCIENTIFIC-METHOD	Agent explains to the user and suggests they use the scientific method while diagnosing the mysterious illness.
INTRODUCE-WORKSHEET	Agent explains usage of the diagnosis worksheet to help the user formulate and refine their hypothesis.
EXAMINE-PATIENT-SYMPTOMS	Agent and user work together to examine symptoms of each of the patients.
UPDATE-WORKSHEET	Agent reminds the user to update the diagnosis worksheet with new knowledge and hypothesis.
READ-DISEASE-BOOKS	Agent guides the user to read relevant disease information in the library, which helps them refine their hypothesis.
INTRODUCE-HEADACHE*	Agent triggers an action resulting in a patient moaning and complaining about having a headache.
TEST-CAMP-ITEMS	User and agent test food items the expedition team took with them from camp.
TEST-OUTSIDE-CAMP-ITEMS	User and agent test food items the team found during their expedition.
TEST-CONTAMINATED-BANANAS	User and agent test the bananas, which end up being contaminated.
INTRODUCE-DIRTY-WATER*	Agent triggers an event causing a door to open and a water bottle to appear in the infirmary room.
INTRODUCE-LEG-CRAMPS*	Agent triggers an event causing one of the patients to complain about leg cramps.
COMPLETE-WORKSHEET	Agent asks user to update all remaining information that has not been entered and formulate their final hypothesis.
REPORT-RESOLUTION	Agent asks user to explain their final hypothesis and how they arrived at their conclusion using the scientific method.
END-SESSION*	Agent thanks user and tells her that the patients will be treated based on her finding.

\*Decisions employed in *Base Model* condition

To examine whether the two conditions differed in terms of participants who successfully solved the mystery, a chi-square test was performed. The results showed that the correlation was significant, (likelihood ratio,  $\chi^2 = 9.14$ , Pearson,  $\chi^2 = 8.84$ ,  $p < .01$ ), indicating that the number of participants who solved the mystery varied significantly between the conditions. Participants in the *Machine-Learned Model* condition solved the mystery at a rate of 92.7%, whereas participants in the *Base Model* condition solved the mystery at a much lower rate of 70.2%.

We also examined differences in time taken by participants to solve CRYSTAL ISLAND's mystery narrative. An ANOVA revealed that differences between the two conditions were statistically significant,  $F(1, 82) = 27.01$ ,  $p < .001$ . These results indicate that participants in the *Machine-Learned*

*Model* condition solved the mystery significantly more efficiently than participants in the *Base Model* condition.

### VIII. DISCUSSION

Empirical evaluation of the DBN-based director agent models supports the promise of data-driven approaches for creating interactive narrative systems. Findings from a controlled experiment with human participants revealed that machine-learned dynamic Bayesian networks significantly outperformed baseline approaches for modeling director agent strategies in a run-time narrative-centered learning environment. Participants who interacted with machine-learned director agent models achieved significantly greater learning outcomes than participants in the baseline condition, and they also solved the science mystery scenario more frequently and efficiently. These experiential measures are highly relevant for educational applications of interactive narratives, which is a key focus of the CRYSTAL ISLAND game environment. A possible explanation for these results is the DBN-based director agent delivered hints and prompts effectively to reduce obstacles to student learning, as well as maintained effective pacing to minimize student boredom and frustration. In contrast, the base model did not provide comparable story-embedded educational support, either in terms of timeliness or content. It should be noted that these results merit complementary analyses that focus on aesthetic dimensions of player experiences. Studies of players' subjective experiences, as well as empirical patterns in their aesthetic experiences, are important for understanding the aggregate impacts of director agent models.

Furthermore, statistical analyses involving leave-one-out cross validation found that dynamic Bayesian networks are capable of accurately modeling human demonstrations of directorial strategies. We believe that DBNs' ability to accurately model human directorial strategies is a key reason for their effectiveness in the run-time CRYSTAL ISLAND environment. The findings are also notable because the DBN models did not incorporate any features from the spoken natural language dialogues that occurred in the Wizard-of-Oz study. While natural language dialogue is widely believed to be a key channel in learning-related communication [28], in this case it was not necessary to induce models of directorial strategies with high degrees of accuracy, or induce a director agent model that effectively enhanced student learning and problem-solving outcomes in a run-time environment.

While supervised machine learning techniques based on dynamic Bayesian networks show considerable promise, they do have limitations. The computational complexity of probabilistic belief updates increases as growing numbers of narrative features or actions are introduced to a model, or as a network's connectedness increases. Increasing the dimensionality of narrative representations can lead to increased predictive power, but these gains come at the cost of runtime performance. A careful balance must be maintained to ensure that director agent models can effectively reason about key aspects of an interactive narrative while still completing inferences within reasonable durations of time. These

TABLE IV  
LEARNING GAINS BY CONDITION

Conditions	Gain Avg.	SD	$t$	$p$
Machine-learned	1.28	2.66	2.03	< 0.05
Base	0.89	3.12	1.23	0.22

TABLE V  
MYSTERY-SOLVING EFFICIENCY BY CONDITION

Conditions	Solved Mystery	Completion Time (s)	
		Mean	SD
Machine-learned	92.7 %	1724	417.0
Base	70.2 %	2229	461.6

tradeoffs are particularly salient in graphically intensive games, which may have limited CPU and memory resources available for AI-related computations.

Furthermore, the current framework's requirement that a corpus collection version of an interactive narrative mirror the final version of the interactive narrative may prove limiting for some categories of games. A growing number of games are updated continuously after release through downloadable content, expansion packs, and in-app purchases. These updates result in game dynamics that frequently change. This relatively recent advancement in game distribution poses notable challenges for data-driven approaches to devising interactive narrative models, but at the same time modern networking technologies introduce substantial opportunities for collecting training data at relatively low costs.

### IX. CONCLUSION AND FUTURE WORK

Computational models of interactive narrative offer considerable potential for creating engaging story-centric game experiences that are dynamically tailored to individual players. Devising effective models of director agent strategies is critical for achieving this vision. We have presented a data-driven framework for machine learning models of director agent strategies using dynamic Bayesian networks. The framework enumerates key considerations of corpus acquisition, model learning, predictive model evaluation, and runtime model evaluation phases of the approach, and its efficacy has been demonstrated through implementations with the CRYSTAL ISLAND interactive narrative environment. Empirical studies involving human participants revealed that dynamic Bayesian networks can accurately predict interactive narrative director agent strategies demonstrated by human players. Furthermore, statistical analyses of data from a controlled experiment indicated that machine-learned DBN director agent models have significant positive impacts on participants' experiential outcomes, including learning and narrative progress. The results suggest that the proposed framework for modeling director agent strategies holds considerable promise for promoting effective and engaging narrative-centered learning experiences.

Several directions for future work are promising. First, user modeling played a minor role in this work, but it could serve a much more prominent role in enhancing computational models of interactive narrative. By maintaining fine-grained models of players' knowledge during interactive narratives, opportunities for new classes of directorial actions may be introduced. Another promising direction is exploring natural language dialogue in interactive narrative. Endowing virtual characters with sophisticated natural language dialogue capabilities offers a mechanism for guiding players through interactive stories using naturalistic and engaging interfaces. During the Wizard-of-Oz study described in this paper, wizards used natural language dialogue to guide participants whenever unexpected behaviors were encountered. While this was emulated through voice-acted speech in the evaluation study, and lack of natural language dialogue did not prevent the machine-learned director agent model from effectively shaping students' narrative-centered learning outcomes, devising adaptive models of interactive dialogue represents a promising line of investigation for future enhancements to narrative-centered learning environments.

#### ACKNOWLEDGMENTS

The authors wish to thank Valve Corporation for access to the Source™ SDK, as well as Joe Grafsgaard, Kate Lester, Megan Mott, and Jen Sabourin for assisting with studies.

#### REFERENCES

- [1] M. Mateas and A. Stern, "Structuring content in the Façade interactive drama architecture," in *Proceedings of the 1st International Conference on Artificial Intelligence and Interactive Digital Entertainment*, 2005, pp. 93–98.
- [2] J. McCoy, M. Treanor, B. Samuel, N. Wardrip-Fruin, and M. Mateas, "Comme il Faut: A system for authoring playable social models," in *Proceedings of the 7th International Conference on Artificial Intelligence and Interactive Digital Entertainment*, 2011, pp. 158–163.
- [3] M. Si, S. Marsella, and D. Pynadath, "Directorial control in a decision-theoretic framework for interactive narrative," in *Proceedings of the 2nd International Conference on Interactive Digital Storytelling*, 2009, pp. 221–233.
- [4] H. Yu and M. O. Riedl, "A sequential recommendation approach for interactive personalized story generation," in *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems*, 2012, pp. 71–78.
- [5] J. Porteous, J. Teutenberg, F. Charles, and M. Cavazza, "Controlling narrative time in interactive storytelling," in *Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems*, 2011, pp. 449–456.
- [6] B. W. Mott and J. C. Lester, "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*, 2006, pp. 977–984.
- [7] D. L. Roberts, M. J. Nelson, C. L. Isbell, M. Mateas, and M. L. Littman, "Targeting specific distributions of trajectories in MDPs," in *Proceedings of the 21st National Conference on Artificial Intelligence*, 2006, pp. 1213–1218.
- [8] M. O. Riedl, C. J. Saretto, and R. M. Young, "Managing interaction between users and agents in a multi-agent storytelling environment," in *Proceedings of the 2nd International Conference on Autonomous Agents and Multiagent Systems*, 2003, pp. 741–748.
- [9] M. Sharma, S. Ontanón, C. Strong, M. Mehta, and A. Ram, "Towards player preference modeling for drama management in interactive stories," in *Proceedings of the 20th International Florida Artificial Intelligence Research Society Conference*, 2007, pp. 571–576.
- [10] D. Thue, V. Bulitko, M. Spetch, and E. Wasylishen, "Interactive storytelling: A player modelling approach," in *Proceedings of the 3rd Artificial Intelligence and Interactive Digital Entertainment Conference*, 2007, pp. 43–48.
- [11] J. Niehaus, B. Li, and M. O. Riedl, "Automated scenario adaptation in support of intelligent tutoring systems," in *Proceedings of the 24th Conference of the Florida Artificial Intelligence Research Society*, 2011, pp. 531–536.
- [12] N. Vannini, S. Enz, M. Sapouna, D. Wolke, S. Watson, S. Woods, K. Dautenhahn, L. Hall, A. Paiva, E. André, R. Aylett, and W. Schneider, "FearNot!: A computer-based anti-bullying-programme designed to foster peer intervention," *European Journal of Psychology of Education*, vol. 26, no. 1, pp. 21–44, Jun. 2011.
- [13] B. Nelson and D. Ketelhut, "Exploring embedded guidance and self-efficacy in educational multi-user virtual environments," *International Journal of Computer-Supported Collaborative Learning*, vol. 3, no. 4, pp. 413–427, 2008.
- [14] W. L. Johnson, "Serious use of a serious game for language learning," *International Journal of Artificial Intelligence in Education*, vol. 20, no. 2, pp. 175–195, 2010.
- [15] J. Thomas and R. M. Young, "Annie: Automated generation of adaptive learner guidance for fun serious games," *IEEE Transactions on Learning Technologies*, vol. 3, no. 4, pp. 329–343, 2011.
- [16] J. P. Rowe, L. R. Shores, B. W. Mott, and J. C. Lester, "Integrating learning, problem solving, and engagement in narrative-centered learning environments," *International Journal of Artificial Intelligence in Education*, vol. 21, no. 1–2, pp. 115–133, 2011.
- [17] B. Li, S. Lee-urban, and M. O. Riedl, "Toward autonomous crowd-powered creation of interactive narratives," in *Proceedings of the 5th AAAI Workshop on Intelligent Narrative Technologies*, 2012.
- [18] R. Swanson and A. Gordon, "Say Anything: A massively collaborative open domain story writing companion," in *Proceedings of the 1st Joint International Conference on Interactive Digital Storytelling*, 2008, pp. 32–40.
- [19] H. Barber and D. Kudenko, "Dynamic generation of dilemma-based interactive narratives," in *Proceedings of the 3rd Artificial Intelligence and Interactive Digital Entertainment Conference*, 2007, pp. 2–7.
- [20] P. Weyhrauch, "Guiding interactive drama," Ph.D. dissertation, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, 1997.
- [21] C. R. Fairclough, "Story games and the OPIATE system: Using case-based planning for structuring plots with an expert story director agent and enacting them in a socially situated game world," Ph.D. dissertation, Dept. of Computer Science, University of Dublin, Trinity College, Dublin, Ireland, 2004.
- [22] M. J. Nelson, D. L. Roberts, C. L. Isbell, and M. Mateas, "Reinforcement learning for declarative optimization-based drama management," in *Proceedings of the 5th International Joint Conference on Autonomous Agents and Multiagent Systems*, 2006, pp. 775–782.
- [23] J. Orkin, "Collective artificial intelligence: Simulated role-playing from crowdsourced data," Ph.D. dissertation, MIT Media Lab, Massachusetts Institute of Technology, Boston, MA, 2013.
- [24] T. Dean and K. Kanazawa, "A model for reasoning about persistence and causation," *Computational Intelligence*, vol. 5, no. 3, pp. 142–150, 1990.
- [25] M. J. Druzdzel, "SMILE: Structural modeling, inference and learning engine and GeNIe: A development environment for graphical decision-theoretic models," in *Proceedings of the 16th National Conference on Artificial Intelligence*, 1999, pp. 342–343.
- [26] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, March 2003, pp. 1157–1182, 2003.
- [27] Y. Yang and X. Liu, "A re-examination of text categorization methods," in *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1999, pp. 42–49.
- [28] D. Litman and K. Forbes-Riley, "Predicting student emotions in computer-human tutoring dialogues," in *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, 2004, Article 351.