Interactive narrative environments offer significant potential for creating engaging narrative experiences. Increasingly, applications in education, training, and entertainment are leveraging narrative to create rich interactive experiences in virtual storyworlds. A key challenge posed by these environments is devising accurate models of director agents’ strategies that make director intervention and action decisions to craft customized story experiences for users. Director agents work behind the scenes to direct a cast of non-player characters and storyworld events for the unfolding narrative. Although a growing body of research has investigated techniques for modeling director agents in interactive narrative, prior work has focused on models learned from simulated data or pre-authored models. A promising approach is developing an empirically driven model of director agents’ decision-making strategies.

In this work, we propose a dynamic Bayesian network framework for modeling director agent narrative decision-making. To create empirically informed models of director agent decision-making strategies, we conducted a Wizard-of-Oz study with an interactive narrative-centered learning environment. In the study, the wizard served as a “human director agent.” Machine learning was used to automatically acquire the conditional probabilities for the dynamic Bayesian networks. The machine-learned models were then empirically evaluated to investigate their effectiveness and efficiency in real-time. Results of the study are encouraging and suggest that empirically driven models of director agent decision-making strategies can offer significant predictive power.
Modeling Director Agents’ Decision-Making Strategies in
Guided Discovery Learning Environments

by
Seung Y. Lee

A dissertation submitted to the Graduate Faculty of
North Carolina State University
In partial fulfillment of the
Requirements for the degree of
Doctor of Philosophy

Computer Science

Raleigh, North Carolina

2012

APPROVED BY:

___________________________________
PATRICK J. FITZGERALD

___________________________________
DAVID L. ROBERTS

___________________________________
R. MICHAEL YOUNG

___________________________________
JAMES C. LESTER
CHAIR OF ADVISORY COMMITTEE
Dedication

This work is dedicated to my family, whose patience, support, and encouragement made it possible.
Seung Lee was born in Busan, South Korea. He immigrated to the United State with his family in 1987. He lived in Virginia at first, but soon after he moved to North Carolina where he attended high school.

After graduating from high school he attended North Carolina State University to pursue a Bachelor of Science degree in Chemical Engineering. In his sophomore year, he added Pulp and Paper Science Technology as a second major. He received his undergraduate degrees in 1997. After working in industry for a few years, he decided to pursue a new career in computer science and began his graduate career in the Department of Computer Science at North Carolina State University. He joined the Intellimedia Initiative while studying in Computer Science, and his M.S. research focused on intelligent pedagogical animated agents for immersive learning environments for the domain of high school physics.

In 2000, he joined a venture-backed enterprise software startup, LiveWire Logic, with colleagues from North Carolina State University. At LiveWire Logic he served as a senior software engineer to develop an automated customer service solution based on computational linguistics technologies. After working at LiveWire Logic more than five years, he returned to the North Carolina State University and entered Ph.D. program in Computer Science under the direction of Dr. James Lester.
Acknowledgments

My appreciation goes out to everyone who helped make this research possible. I would like to express my sincere gratitude to my advisor, James Lester, for his countless hours of guidance and support. I would also like to thank my committee members, Patrick Fitzgerald, David Roberts, and Michael Young, for their direction and collaboration.

My thanks also go to the IntelliMedia Group team members for their contribution to the project. I have benefited greatly from working with a bright and kind group of people including Alok Baikadi, Kristy Boyer, Julius Goth, Joe Grafsgaard, Eun Young Ha, Eleni Lobene, Samuel Leeman-Munk, Chris Mitchell, Jonathan Rowe, Marc Russo, Jennifer Sabourin, Andy Smith, and Robert Taylor. Thanks to the Megan Mott and Kate Lester who helped me on the character voice and research study. I especially thank Bradford Mott for his help and support.

Finally, I thank my family for their patience, support, and encouragement.
Table of Contents

List of Tables .................................................................................................................. vii
List of Figures .................................................................................................................. viii

Chapter 1 Introduction .................................................................................................. 1
  1.1 Motivation .............................................................................................................. 2
  1.2 Approach .............................................................................................................. 3
  1.3 Contributions ....................................................................................................... 4
  1.4 Organization ....................................................................................................... 5

Chapter 2 Background and Related Work .................................................................. 7
  2.1 Offline Narrative Generation ............................................................................... 7
  2.2 Interactive Narrative Generation ......................................................................... 9
  2.3 Narrative-Centered Learning Environments ....................................................... 13

Chapter 3 Director Agents for Interactive Narrative Environments ......................... 17
  3.1 Problem Definition ............................................................................................. 17
  3.2 Proposed Approach ........................................................................................... 19
  3.3 Architecture ........................................................................................................ 20

Chapter 4 Corpus Collection ....................................................................................... 22
  4.1 CRYSTAL ISLAND ............................................................................................. 22
  4.2 CRYSTAL ISLAND: Wizard-of-Oz Version ..................................................... 23
  Wizard-of-Oz Functionality ....................................................................................... 26
  4.3 Example Scenario .............................................................................................. 27
  4.4 Narrative Observations ....................................................................................... 28
    4.4.1 Narrative Action Decisions ............................................................................ 30
    4.4.2 Narrative Progress ....................................................................................... 30
    4.4.3 Physical State ............................................................................................... 31
  4.5 Corpus Collection Procedure .............................................................................. 32
    4.5.1 Participants .................................................................................................. 32
    4.5.2 Participant Procedure .................................................................................. 33
    4.5.3 Wizard Protocol .......................................................................................... 33

Chapter 5 Modeling Director Agents with Dynamic Bayesian Networks .................. 35
  5.1 Feature Selection .................................................................................................. 36
  5.2 Intervention Decision Strategies ......................................................................... 37
  5.3 Action Decision Strategies .................................................................................. 39
  5.4 Integrated Real-Time Models .............................................................................. 40

Chapter 6 Evaluation ................................................................................................... 44
  6.1 Hypotheses .......................................................................................................... 44
  6.2 Predictive Accuracy Evaluation ............................................................................ 45
    6.2.1 Intervention Decision Model ........................................................................ 45
    6.2.2 Action Decision Model ................................................................................. 47
Appendix D Student pre/post-test ................................................................. 104

Appendix C Description: A Dynamic Bayesian Network Model for Narrative Decision-Making .......... 98
C.1 Model Structure ...................................................................................... 100
C.2 Inference .................................................................................................. 102
C.2 Learning ................................................................................................. 102

Appendix B Wizard Protocol ................................................................. 88
B.1 Wizard Interaction Guideline ................................................................. 88
B.2 Narrative Decisions Guideline .............................................................. 92
B.3 Character Symptoms and Object Test Results Guideline ....................... 95
B.4 Disease Symptoms Guideline ............................................................... 96

Appendix A Corpus Collection Handout Materials ........................................ 82
A.1 CRYSTAL ISLAND Storyworld .............................................................. 82
A.2 Interacting with Virtual Environment ..................................................... 86
A.3 General Instructions for Interacting with Virtual Environment ................ 86

References .................................................................................................. 73

Chapter 7 Conclusion ............................................................................... 68
7.1 Summary .............................................................................................. 68
7.2 Limitation ............................................................................................ 70
7.3 Future Work ......................................................................................... 70
7.4 Concluding Remarks ........................................................................... 72

6.4 Discussion ............................................................................................ 65
6.3 Runtime Evaluation ............................................................................... 48
6.3.1 Study Design Considerations .......................................................... 48
6.3.2 Evaluation Design ........................................................................... 50
6.3.3 Method ............................................................................................ 53
6.3.4 Results ............................................................................................ 57
6.3.5 Individual Differences .................................................................... 61

Appendices .............................................................................................. 81

Appendix A Corpus Collection Handout Materials ........................................ 82
List of Tables

Table 1. Narrative action decisions ........................................................................................................... 29
Table 2. Narrative arc phases ...................................................................................................................... 31
Table 3. Infirmary and laboratory locations ............................................................................................... 32
Table 4. Classification results of director agent intervention models ..................................................... 46
Table 5. Classification results of director agent action models ............................................................... 47
Table 6. Minimal set of director action decisions ...................................................................................... 50
Table 7. Set of director action decisions as islands .................................................................................... 51
Table 8. Full director action decisions ..................................................................................................... 52
Table 9. Demographics in each condition .................................................................................................. 53
Table 10. Learning gains and t-test statistics ............................................................................................ 57
Table 11. In-game problem-solving performances .................................................................................... 58
Table 12. Means and SD on PQ raw scores ............................................................................................... 59
Table 13. Means and SD on IMI raw scores .............................................................................................. 60
Table 14. Total time on development ......................................................................................................... 61
Table 15. Gender differences in learning gains and t-test statistics .......................................................... 62
Table 16. Gender differences in-game problem-solving performances ................................................... 63
Table 17. High- and Low-Achievers in-game problem-solving performances differences ...................... 64
Table 18. Game play experiences differences in learning gains and t-test statistics ................................. 65
Table 19. Game play experience differences in-game problem-solving performances ............................. 65
List of Figures

Figure 1. Data-driven narrative methodology ........................................................................4
Figure 2. Interactive Narrative Architecture ........................................................................20
Figure 3. WOZ-enabled CRYSTAL ISLAND ......................................................................23
Figure 4. CRYSTAL ISLAND from nurse’s perspective .........................................................25
Figure 5. Plot Graph for Crystal Island: Wizard-of-Oz Version ............................................26
Figure 6. WOZ-enabled CRYSTAL ISLAND ......................................................................27
Figure 7. A prototypical narrative arc ..................................................................................31
Figure 8. Dynamic Bayesian network model for intervention decision-making strategies ................37
Figure 9. Dynamic Bayesian network model for narrative action decision strategies ..................40
Figure 10. CRYSTAL ISLAND story-based learning environment with integrated models ..............41
Figure 11. Director Agent driven CRYSTAL ISLAND story-based learning environment ...... 42
Figure 12. CRYSTAL ISLAND real-time interaction ............................................................43
Figure 13. Study Procedure ................................................................................................56
Chapter 1

Introduction

Recent years have witnessed significant growth in research on interactive narrative environments that create engaging narrative experiences for education, training, and entertainment applications (Aylett et al. 2005; Johnson & Wu 2008; Mateas & Stern 2003; Rowe et al. 2010). Utilizing the inherent structure of narrative, interactive narrative planning offers significant potential for creating engaging and effective narrative experiences (Barber & Kudenko 2007; Cavazza, Charles, & Mead 2002; Nelson & Mateas 2005). A broad range of computational models of interactive narrative planning has been investigated to build coherent narrative structures while integrating users’ interactions in real-time (Magerko & Laird 2003; Mott & Lester 2006; Nelson et al. 2006; Roberts et al. 2006). A common metaphor these models share is employing a director agent, which is a drama manager that works behind the scenes to direct a cast of non-player characters and storyworld events for the unfolding narrative (Bates 1992; Weyhrauch 1997). Through their narrative actions, director agents subtly guide users through an intended narrative experience.
1.1 Motivation
Interactive narrative-centered learning environments can form the basis for discovery learning that supports users’ active exploration of a subject matter (Bruner 1961). Discovery learning encourages users to learn by trial-and-error. Utilizing the scientific method, users pose questions, design and perform experiments, collect data, and evaluate hypotheses (de Jong & Joolingen 1998). Despite the potential benefit of discovery learning, studies have indicated that it can be ineffective when users receive no guidance in the form of coaching and hints from a teacher or learning environment (Mayer 2004, Kirschner, Sweller, & Clark 2006). These studies suggest that discovery learning that is accompanied by guidance can be more effective than pure discovery learning (de Jong & Joolingen 1998; Shulman & Keisler 1966).

Interactive narrative-centered learning environments actively monitor users interacting with the unfolding storyworld to make decisions regarding the next action to perform in service of guiding users’ learning experiences. Through this process, these environments attempt to make effective narrative decisions while managing the story structure and scaffolding user interaction. To guide users through an intended narrative experience these environments employ a director agent. Director agents actively observe the unfolding storyworld events and determine when it is most appropriate to intervene with the next director agent action to perform, such as causing non-player characters to talk with users. A key challenge posed by these environments is devising accurate models of director agent strategies that determine narrative intervention and action decisions to craft customized story experiences for users.

Prior approaches have utilized either handcrafted models of decision strategies (Weyhrauch 1997; Mateas and Stern 2003) or have learned models based on simulated users (Roberts et al. 2006). A promising approach to building an effective director agent decision strategies model for interactive narrative environments is adopting an empirically driven method. By utilizing a corpus of collected human interactions within a narrative environment, models of director agent decision-making strategies can be learned from data.
1.2 Approach

To create an accurate empirically informed computational model of director agents, a number of requirements must be satisfied. First, the model should support making effective narrative decisions about director agent actions utilizing a data-driven approach. Second, the model should not only indicate the best narrative decision to make but also the appropriate time to intervene. Third, the model should be able to integrate observations from a number of storyworld events while making decision strategies. Fourth, the model must be capable of operating in real-time to support interactivity.

To address these requirements, we propose a dynamic Bayesian network (DBN) approach to modeling director agent decision strategies. The approach supports learning models from a corpus, integrating different sources of evidence affecting decision-making strategies, creating both narrative intervention and action decision models, and the capability of operating in real-time. Our research methodology involves four stages. Figure 1 depicts the first three stages of our methodology for creating a DBN based data-driven narrative decision-making models.

- **Corpus Acquisition:** To collect a corpus of human interactions within a narrative environment, a Wizard-of-Oz (Dahlbäck, Jänsson, and Ahrenberg 1993) study using a narrative-centered learning environment within a 3D virtual storyworld is conducted. In the study, the wizard acts as the director agent that provides data for learning models of decision strategies.

- **Model Building:** Dynamic Bayesian networks are used to model narrative decision-making. Because interactive narrative is a time-based phenomenon, DBNs provide a natural representation for describing worlds that change dynamically over time. The models are built on the corpus collected from the Wizard-of-Oz study.

---

1 A Wizard-of-Oz study is a research methodology in which users interact with what might appear as an autonomous computer system, but the computer is actually driven by a person (the “wizard”) in another location.
• **Model Learning**: Machine learning is then used to automatically acquire the conditional probabilities for the dynamic Bayesian networks.

• **Runtime Model Evaluation**: The learned model of director agent decision-making strategies can then be incorporated in runtime system.

![Figure 1. Data-driven narrative methodology](image)

### 1.3 Contributions

The work reported in this dissertation makes the following contributions, some of which have been reported in the interactive narrative and intelligent tutoring systems literatures:

**Contribution 1**: To investigate the impact of users’ narrative experiences on learning outcomes, a Wizard-of-Oz (WOZ) study was conducted with middle school students interacting with a narrative-centered learning environment for microbiology. With narrative
planning, tutorial planning, and natural language dialogue functionalities provided by wizards, the WOZ study revealed that in interactive story-based learning, students exhibited positive learning outcomes (Lee, Mott, & Lester 2010a; Lee, Mott, & Lester 2010b).

**Contribution 2**: A dynamic Bayesian network was designed and developed to make director agent narrative decisions. The network parameters were learned from a corpus collected in a Wizard-of-Oz study. The performance of the resulting model was evaluated with respect to predictive accuracy and yielded encouraging results (Lee, Mott, & Lester 2011a).

**Contribution 3**: A dynamic Bayesian network framework was designed to model empirically informed director agent intervention decision strategies by utilizing collected corpus from a Wizard-of-Oz study conducted in narrative-centered learning environment. The performance of the resulting model was evaluated with respect to classification accuracy and produced promising results (Lee, Mott, & Lester 2011c).

**Contribution 4**: A modeling methodology was created to support the discovery of an empirically driven director agent decision-making strategies using dynamic Bayesian network structures in interactive narrative environments (Lee, Mott, & Lester 2011b).

**Contribution 5**: A director agent trained on observations of a “human drama manager” has been empirically demonstrated to increase learning gains for users interacting with it in a narrative-centered learning environment (Lee, Mott, & Lester 2012).

**Contribution 6**: A director agent trained on observations of a “human drama manager” has been empirically demonstrated to increase user in-game performance for users interacting with it in a narrative-centered learning environment (Lee, Mott, & Lester 2012).

**1.4 Organization**
The organization of dissertation is structured as follows. In Chapter 2 we summarize related work on intelligent narrative technologies and narrative-centered learning environments.
Various approaches that have been previously taken in narrative generation are described. In Chapter 3, we describe the director agents in interactive narrative and the problem they faced. We provide our proposed approach and general architecture for director agents for interactive narrative environments. The Wizard-of-Oz corpus collection, its design, and procedure are described in Chapter 4. Chapter 5 presents the use of dynamic Bayesian network models for director agent decision-making strategies. Chapter 6 discusses the process of integrating the machine-learned models of director agent decision-making strategies into the real-time system. Chapter 7 presents the methods and results from the evaluations conducted on the director agent decision-making strategies in guided discovery learning environment. Details of study hypotheses, evaluation methodology, and their results are discussed. Chapter 8 provides a summary of the key contributions and promising future directions for future work.
Chapter 2

Background and Related Work

Narrative technologies have been the focus of a number of research efforts. In this chapter we review work on both offline narrative generation and interactive narrative generation.

2.1 Offline Narrative Generation

TALE-SPIN is a story generation system that creates Aesop’s fable-like stories by simulating behaviors of characters in a storyworld (Meehan 1977). The inference mechanism takes an asserted event and generates the consequence of the event. A designer specifies one or more characters and initial settings for TALE-SPIN. The designer also provides a character’s goal to achieve during story creation. TALE-SPIN reasons about the given characters based on its representation of the world and attempts to satisfy its goal. TALE-SPIN consists of three active components. Given a goal, the problem solver decomposes it into sub-goals. The assertion mechanism takes an event and adds it into the world model.

UNIVERSE is a story generation system that uses a planning approach to generate open-ended soap-opera-style stories (Lebowitz 1984; Lebowitz 1985). It operates in two primary phases: character creation and planning. Universe defines person frames for each character in the story, where it includes information about a character’s name, trait, goals, stereotypes, and
status (Lebowitz 1984). The planner produces stories by planning over the *plot fragments*, which includes information about characters, constraints, goals, and sub-goals (Lebowitz 1985). The planner creates a plan by selecting appropriate plot fragments to achieve a various author-specific goals. To determine the involvement of a character, the planner identifies whether the character’s *person frame* matches *plot fragment*’s constraints and the role.

**Minstrel** is a story generation system in the domain of King Author and his Knights. It employs case-based reasoning to adapt a model for creativity (Turner 1994). In **Minstrel**, narrative creation is driven by a set of author level goals for each narrative. The author goals are categorized into thematic, drama, consistency, and presentation goals. Thematic goals denote the theme that will be exhibited by the story. A drama goal defines events that increase elements such as suspense, tragedy, foreshadowing, and characterization. A consistency goal entails satisfying the expectation of the actions in the storyworld, how the actions should be achieved in the storyworld without featuring inconsistency. A presentation goal defines how events in the storyworld should be ordered and provides proper presentation.

**Fabulist** is a plan-based narrative generation system (Riedl & Young 2004). The system identified two properties for successful construction of a story: plot coherence and character believability. Plot coherence requires that the storyworld events are perceived to be meaningful and relevant to the outcome of the story. Character believability requires that the character actions that are perceived to be motivated by a character’s internal beliefs and desires. It uses the intent-driven partial order causal link (IPOCL) narrative planner to ensure that character actions are believable and incorporate a model of character intention recognition. An intention recognition component determines whether character actions added to the plan are intentional. Character intention is caused by some motivating storyworld events.
**PREVOYANT** is a plan-based approach to create a surprise outcome in story by generating flashback and foreshadowing (Bae 2008). Flashback is an instance of analepsis that provides a backstory to explain what cause the outcome. Foreshadowing is an instance of prolepsis that provides an implicit hint on causes of the outcome. Given a chronologically ordered set of input story events, **PREVOYANT** reconstructs story events to create a surprise outcome. **PREVOYANT** identifies a set of Significant Events (SEs), which may become surprising events. SEs are the events that are connected directly to the goal state. For each SE, a set of Initiating Events (IEs) is identified. An IE is a set of causal chains that lead from the initial state to the SE. For the SE that has connections for separable causal IE chains that do not affect its links to other plans, the chain can be removed and the SE is surprising event. One of the separable IE elements will be used to foreshadow the SE and the entire IE is used for providing a flashback for the SE.

### 2.2 Interactive Narrative Generation

The **OZ PROJECT** (Bates 1992; Mateas 1997) hypothesized that to generate engaging experiences with computer-assisted interactive drama, the user must be able to “suspend disbelief” so that they can imagine the computer simulated world is in fact real. In order to simulate users’ responses to an interactive drama in virtual environment, the group conducted an experiment with human actors. The experiment revealed that even though there is some obvious inconsistent behavior in characters, it was tolerable and does not disturb the “suspension of disbelief” (Kelso, Weyhrauch, & Bates 1993). Thus, a significant research effort had focused on character believability (Loyall 1997). However, there are limitations on agents’ behaviors. They were unable to reason about the long-term goals and the behaviors that lead to meaningful conclusions. The drama manager was introduced in order to provide direction to the agents in the storyworld. The drama manager uses a plot graph to model interactive narrative plot. A plot graph is based on a directed acyclic graph (DAG), where the nodes are major events or situations and the arc decides partial or temporal ordering between the nodes. The drama manager maintains active nodes in the plot graph that can be readily
achieved. By continuously monitoring the behaviors in characters and user, the drama manager determines character actions and changes in the storyworld to ensure the story progression.

**I-STORYTELLING** is an interactive storytelling system that uses autonomous characters-based approaches to generate stories (Cavazza, Charles, & Mead 2002). In I-STORYTELLING, the user is a spectator in the sit-com-like interactive storytelling system, where the user can influence the story by intervening in the story. The user can provide spoken advice to the characters and interact with objects in the world. The characters’ roles are generated from hierarchical task network (HTN) planning. An HTN planner is used to generate plans achieving characters’ goals. Each character in the storyworld generates a plan by searching through the HTN, allowing the character to interleave planning and execution. The lowest-level operators of each plan are executed within the environment in the form of Unreal Tournament actions and the action outcome is then passed back to the HTN planner. However, if there is a physical intervention in the world by the user, which in turn causes a character’s action to fail, then this action failure will trigger re-planning.

**MIMESIS** is a plan-based system for interactive narrative (Young 2001; Young & Riedl 2003). It is a two-tier architecture; the Mimesis Unreal Tournament Server (MUTS) and the Memesis Controller (MC). The MC reasons about high-level narrative structure and user interactions. It is responsible for generating the story and maintaining coherent narrative experiences. Two primary components in the MC that serve to create the content of story are the narrative and discourse planner. The narrative planner determines all actions that occur in the environment as story unfolds. It searches for a sequence of actions (story plan) to be carried out in the story. The discourse planner selects communicative mechanisms that can be used in the game engine to convey the unfolding story actions to the user. Both the narrative and discourse planner use the Longbow planning system (Young, Pollack, & Moore 1994), a hierarchical partial-order causal link planner. The MUTS manages low-level user interactions in the game environment. Once story and discourse plans are generated, information is
transmitted to the MUTS. Upon receiving the messages, the MUTS builds a directed acyclic graph (DAG), where nodes represent individual actions and arcs define temporal constraints between the nodes. Nodes in DAG are one-to-one mapped to game engine functions.

**FAÇADE** is an interactive drama manager that addresses the balance of character behaviors and plot creation in a dialogue-oriented interactive story (Mateas 2002; Mateas & Stern 2003; Mateas & Stern 2005). In FAÇADE, the user plays the role of a guest character visiting her friends, Grace and Trip, and learns that they have a serious marital problem. The user can communicate with characters via natural language dialogue by typing sentences and navigate around an apartment. The user’s interactions influence the course of the unfolding drama. FAÇADE uses fine-grained plot element called beats. A **beat** is the smallest element of action in the story that can move the story forward. Beats are implemented in a reactive behavior language, **ABL** (Mateas 2002). ABL is an extension of Loyall’s **HAP** system (Loyall 1997). The drama manager sequences beats that encode character goals and the current story state so that the characters appear motivated by their own goals and desires while the system is creating a strong story arc.

**U-DIRECTOR** is a decision-theoretic approach of narrative planning architecture for storytelling environments (Mott & Lester 2006). The primary objective of U-DIRECTOR is to cope with uncertainty associated with the tasks in narrative environments. Thus, the goal of the system is to satisfy narrative rationality, which is to maximize expected narrative utility by reasoning under uncertainty in a principled manner about narrative objectives, storyworld state, and user state. To address this objective U-DIRECTOR uses a dynamic decision network architecture for the director agent to achieve narrative rationality. To define narrative decision cycles, U-DIRECTOR defines three distinct time slices. Each time slice encodes three principle knowledge sources to represent the probabilistic beliefs of the director agent. They are **narrative objectives** (plot progress and narrative flow), **storyworld state** (plot focus and physical state), and **user state** (user goals and belief and user experiential state). U-DIRECTOR was implemented and evaluated in **CRYSTAL ISLAND**, an interactive narrative-centered
learning environment. The network contains approximately 200 chance nodes and over 400 links representing three principled knowledge sources; the sub-networks encode over 7000 conditional probabilities; the narrative utility network is structured with over 50 utility nodes.

**Search-Based Drama Management (SBDM)** was developed by Weyhrauch (1997). SBDM is based on an abstracting content of the story into a set of plot point, choosing a set of drama manager (DM) actions, and specifying an evaluation function. Plot points represent significant events in the story. They are encoded in a directed acyclic graph (DAG) with the edges specifying ordering. Sequences of plot points in the DAG can be considered a valid story. DM actions reconfigure the storyworld in an effort to guide users through an intended narrative. DM actions include things such as *permanent deniers* (prevent plot point ever happening), *temporary denier* (suspend plot point until *reenabler* action happens), *causer* (make plot point happen), and *hint* (make plot point likely to happen). An evaluation function encodes an author’s aesthetic. It considers a sequence of plot points and the history of DM actions, and guides story towards one that has a higher evaluation score. Weyhrauch (1997) reported strong results of using SAS+, Sampling Adversary Search that is a modification of mini-max game-tree search algorithm to optimize the evaluation function, in *Tea for Three* interactive fiction. Nelson and Mateas (2005) explored SAS+ in the *Anchorhead* story in simulated experiments. However, they were not able to reproduce the same results as Weyhrauch. They encountered difficulties in choosing the set of DM actions was ineffective.

**Declarative Optimization-Based Drama Management (DODM)** continued work on SBDM (Nelson & Mateas 2006) that seeks to optimize the use of a given set of plot points, drama manager actions, and an author’s evaluation function with a reinforcement learning algorithm is used to learn policies. For reinforcement learning, the current history of the sequence of plot points is defined as a state, DM actions and null actions (do nothing) are specified as a set of actions, the player model is defined as transition, and the evaluation function is a reward function. Although they are well defined, the nature of reinforcement learning poses significant challenges. It is highly stochastic and the state space is very large.
To remedy the problem, Nelson and Mateas (2006) presented temporal-difference (TD) learning (Sutton 1988) methods and self-adversarial/self-cooperative exploration (SASCE) methods. They showed that both approaches outperformed search-based methods.

2.3 Narrative-Centered Learning Environments
Narrative-centered learning environments provide users with the ability to actively participate in problem-solving activities by leveraging narrative to create engaging experiences in rich virtual interactive storyworlds. A broad range of techniques has been proposed to create interactive story-based learning environments that are both engaging and pedagogically effective. In this section we review the most relevant projects that employ story-centric problem-solving in interactive narrative-centered learning environments.

**Teatrix** is an interactive learning environment designed to help users in the process of a collaborative fairy tale based story creation (Prada, Machado, & Paiva 2000). **Teatrix** allows users to choose their own characters and interact with one another in a distributed 3D environment. The user can create a story initially by defining its scenes, characters, and items that can support character performance. Once the initial story creation is complete, users play the story as one of the characters. Characters that are not controlled by the users act autonomously based on their role in a goal-directed manner in the story. To ensure that the story is coherent, **Teatrix** uses a director agent to control the narrative of the story.

**Mission Rehearsal Exercise (MRE) Project** is a rich 3D virtual reality training environment for teaching decision-making skills in peacekeeping and disaster relief mission to military personnel (Swartout et al. 2001). To create compelling interactivity in the story, autonomous agents are introduced into the storyworld. Scripted characters perform specific actions that are pre-specified path and behaviors. Autonomous agents’ behaviors are not scripted and are implemented in Soar, a general architecture for building intelligent agents. To accommodate both structured and unstructured interactivity an approach to story creation, a *StoryNet*, is employed. Within a StoryNet, users are given the freedom to take control,
initiate actions, and make choices. The links between the nodes are scripted sequence of events in which the user remains passive.

**Carmen’s Bright IDEAS** (CBI) is an agent-based interactive pedagogical drama. CBI is an interactive health intervention system designed to teach social problem-solving skills to mothers of pediatric cancer patients (Marsella et al. 2003). Interactive stories are organized into three acts. The first act is a back-story that introduces various problems that Carmen is facing, such as her son’s cancer. The second act takes place in an office where Carmen discusses her problems with a clinical counselor Gina, an autonomous agent. CBI uses a social problem-solving method called Bright IDEAS (Varni et al. 1999) to help Carmen find a solution to her problems. The final act portrays the outcome of Carmen’s application of Bright IDEAS. The purpose of CBI is to teach mothers how to apply the Bright IDEAS method when a problematic situation occurs. The learner interacts with the CBI by making choices for Carmen such as what problems she needs to work on and how to cope with the problems she is facing. CBI ensures interactivity with the learner by facilitating problem-solving and dialogue model to achieve goals and using an emotional appraisal model to evaluate agent emotions and a behavior generation model to manager non-verbal activities.

**FEARNOT!** is a storytelling application for an anti-bullying education (Aylett et al. 2005). It uses affectively-driven autonomous agents to dynamically generate dramatic performance for virtual drama. The interaction structure of FEARNOT! first introduces the school and characters and a dramatic episode in which a bullying incident occurs. The victim character asks the user for advice by activating the goal of a help-requesting speech act, *askforhelp*. Then the user tries to suggest a coping behavior. She talks with the victim character via an interaction interface where she types instructive actions for the victim character to do next. When the agent receives the written utterances, they are converted to language actions using a template-based language system. Upon receiving a suggestion, the agent analyzes the environment, acts, and emotional state, and adds a new intention to achieve the active goal.
**Stability and Support Operations (SASO)** is a multi-agent system that features virtual humans with sophisticated models of emotion (Gratch & Marsella 2004), social behavior (Gratch et al. 2007), and multi-modal multi-party dialogue (Traum et al. 2008). It integrates with interactive virtual human technologies for a training environment (Kenny et al. 2007). SASO demonstrates an advanced virtual human technology, where trainees are to communicate and conduct negotiations with an embodied virtual doctor and a village elder to move a clinic to another part of the town.

**Tactical Language And Culture Training System (TLCTS)** is an interactive environment for learning foreign languages and culture (Johnson & Wu 2008). It is designed to assist learners in acquiring spoken language communication skills by providing interactive lessons that provides practice with specific skills through interactive games. The system includes intelligent tutoring functionalities, spoken dialogue, natural language processing, automated speech recognition, and animated agents. TLCTS is the widely used and most popular TLCTS course is the Tactical Iraqi™, which teaches colloquial Iraqi Arabic. A study with the Marine Corps indicates that the Tactical Iraqi™ course is effective. The TLCTS training course includes a Skill Builder, which focuses on communicative tasks. Depending on the type of exercise, the system can provide feedback on pronunciation, word choice, grammatical forms, or cultural pragmatics. The Mission Game is also part of TLCTS training course. The learner communicates with Iraqi non-player characters (NPC) using speech and gesture in order to complete the mission.

Although prior work on interactive narrative and narrative-centered learning environments has investigated techniques for modeling director agents, they are focused on either models learned from simulated data or pre-authored models. Little work has investigated devising empirically informed models of director agent narrative decision-making. Drachen et al. (2009) proposed an approach of designing intelligent drama managers by learning from a corpus data where users and expert human game masters interact in pen-and-paper RPG game sessions. Data collected from game masters, analogous to a director agent, is utilized to
learn drama manager actions while maintaining a coherent narrative. Although the approach is closely related to ours, interactions are limited to pen-and-paper RPG game sessions, and they did not utilize a computational model that utilizes the collected corpus data and any empirical evaluations to support their claims.
Chapter 3

Director Agents for Interactive Narrative Environments

Director agents manage the unfolding stories in interactive narrative environments. In this chapter, we first define the primary objective of modeling director agent decision-making strategies. We then present a novel approach to modeling director agent decision-making strategies that employ a data-driven methodology. We then introduce an architecture for interactive narrative systems in which the proposed director agent model is embedded.

3.1 Problem Definition

Director agents for interactive narrative craft global story arcs utilizing a set of plot points (Bates 1992) representing a partial order of significant story events and a set of director actions that reconfigure the storyworld in an effort to guide users through an intended narrative. To create engaging user experiences, director agents must determine the most effective sequence of plot points and director actions to perform while balancing character believability, plot coherence, and author aesthetics. To sequence a set of plot points, a director agent might employ a search mechanism and an evaluation function encoding key features of the narrative, such as author aesthetics, agency, thought flow, and location flow.
(Nelson & Mateas 2006). Using the evaluation function to rank the potential plot point sequences, the director agent can select the next plot point on which to focus the story. Once selected, the director agent can utilize its set of director actions to guide the story to the selected plot point, e.g., directing a character to perform an action in the environment.

Thus, director agents typically consider three components while crafting a global story arc for interactive narrative. They typically utilize (1) a set of plot points representing significant story events, such as discovering a key clue while solving a mystery, (2) a set of director agent actions that provide a means for guiding users through an intended narrative experience, such as causing non-player characters to talk with users, and (3) an evaluation function to determine the most desirable sequence of plot points to optimize the story quality (Bates 1993; Weyhrauch 1997; Nelson and Mateas 2005). By utilizing these components, director agents create engaging user experiences. In the course of interactive narratives, director agents actively observe the unfolding storyworld events and determine when it is most appropriate to intervene with the next director agent action to perform in service of guiding users through an intended narrative experience.

Through this process, director agents attempt to determine effective decision strategies while managing the overall story structure and plot coherence. A significant challenge posed by interactive narrative environments is devising accurate models of director agents that make effective decision strategies to craft customized story experiences for users. Although a growing body of research has investigated techniques for modeling director agents in interactive narrative, prior work has focused on models learned from simulated data or pre-authored models. A promising approach is developing an empirically driven model of director agents’ decision-making strategies. By utilizing a collected corpus of human interactions within an interactive narrative environment, models of director agents’ decision-making strategies can be learned from data.
3.2 Proposed Approach
To create an accurate empirically driven computational model of director agents, we propose a dynamic Bayesian network (DBN) approach. Interactive narrative is a time-dependent phenomenon. 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 the unfolding story. Dynamic Bayesian networks (DBNs) explicitly characterize models’ belief state over time. DBNs provide a natural representation for describing worlds that change dynamically (Dean & Kanazawa 1989).

For creating a DBN based data-driven director agent decision-making strategies, we conducted a Wizard-of-Oz (WOZ) corpus collection study with thirty-three participants using a narrative-centered learning environment within a 3D virtual storyworld. In the corpus collection, users assumed the role of a medical detective solving a science mystery while the wizard provided director agent functionalities for the system. Throughout the corpus collection, detailed trace data were collected for all wizard decision-making and all navigation and manipulation activities within the virtual environment. The resulting corpus of trace data was utilized to learn models of director agent decision-making strategies using dynamic Bayesian networks. To create the DBN models for the narrative intervention and action decision strategies, we manually engineered the network structure and learned the parameters. Each of the network structures was integrated with different observable variables, and their conditional probabilities were learned using leave-one-out cross validation. The Expectation-Maximization (EM) algorithm was used to learn the conditional probability table (CPT) parameters.

To explore the real-time effectiveness of the machine-learned models of director agent decision-making strategies, the intervention decision model and the action decision model were integrated into the CRYSTAL ISLAND story-based learning environment. The environment is identical to the WOZ corpus collection study environment except non-player
character interactions are driven by the director agent. The evaluation methodology and results of the machine-learned models of director agent decision-making strategies are described in Chapter 7.

### 3.3 Architecture
The interactive narrative architecture utilized in our work is shown in Figure 2. All interactions between the user and the storytelling environment are mediated by the interface manager, which provides functions of rendering and activities within the storyworld. The storytelling environment employs three main sets of resource components: a 3D storyworld, rendering and sound effects, and believable characters. Using these resources the storytelling environment can provide information regarding characters and actions in the world so that director agent can maintain its plot creation and character behavior control. The director agent actively observes all user activities, as well as actions of characters and objects in the storyworld. Using these observations the director agent directs events to craft the narrative.

![Interactive Narrative Architecture](image)

*Figure 2. Interactive Narrative Architecture*
The director agent has access to the narrative observations that occur while user interaction occurs in the storytelling environment to determine the next director agent actions to maintain the coherence of the narrative. The *narrative history* maintains the knowledge of the prior narrative decisions made by the director agent. The *narrative progress* observation models the storyworld’s narrative structure. The *user knowledge* represents the user’s beliefs about the salient facts of the story learned through interactions with the environment and other characters. The *physical state* represents the current location of the characters in the storyworld. The *user actions* indicate the user activities where they are specified by user character’s interactions with objects and characters in the storyworld. The *time index* represents the overall timeline of the storyworld to provide the temporal-based evidence for the director agent decision-making strategies.
Chapter 4

Corpus Collection

To investigate director agent decision-making strategies, a Wizard-of-Oz data collection was conducted with a customized version of the CRYSTAL ISLAND narrative-centered learning environment (Lee, Mott, & Lester 2010b). After introducing CRYSTAL ISLAND, we describe the custom episode created for this data collection along with the Wizard-of-Oz functionalities introduced into the environment.

4.1 CRYSTAL ISLAND

CRYSTAL ISLAND is a narrative-centered learning environment developed for middle school students for the domain of eighth-grade microbiology (Rowe et al. 2011). It is built with Valve Corporation’s Source™ engine, the technology behind Half Life® 2. CRYSTAL ISLAND features a science mystery set on a recently discovered tropical island where a research station has been established to study the island’s unique flora and fauna. Within the story, the user plays the role of the protagonist attempting to discover the identity and source of an infectious disease plaguing the research station. Throughout the mystery, the user is free to explore the world and interact with other characters while forming questions, generating hypotheses, collecting data, and testing hypotheses. The user can pick up and manipulate
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 user completes an in-game diagnosis worksheet to organize her thoughts about the patients’ symptoms, the likelihood of potential diseases (based on their expected symptoms, incubation period, and transmission source), and her final diagnosis. Upon completing the diagnosis worksheet, the user verifies its contents with the camp nurse and develops a treatment plan for the sickened CRYSTAL ISLAND researchers.

4.2 CRYSTAL ISLAND: Wizard-of-Oz Version
For the corpus collection, a custom episode of CRYSTAL ISLAND (Figure 3) was created featuring a companion agent who assists the user in solving the mystery. This episode features six characters Alyx Reid (player), Kim Lee (camp nurse and companion agent), Bryce Reid (lead scientist), Ford Patterson (zoologist), Audrey Newsome (botanist), Quentin Nash (camp cook), and Al Cunningham (camp foreman). The user plays the role of Alyx Reid visiting her father, Bryce, who serves as the research station’s lead scientist.
Alyx has arrived at Crystal Island to visit her father whom she has not seen for awhile. As she approaches the dock, she hears news that her father has fallen ill from Al, the camp foreman. Al 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. Al finds this very unusual since they were known to adhere closely to schedule. Fearful for their safety, Al leads a search team to locate them. After two days of searching, the research team discovers that the expedition team had fallen ill on the south side of the island. It appears the group lost their way, became ill, and could not return to camp. They are in the infirmary and are being attended to by the camp’s nurse. Upon hearing the news, Alyx goes to the infirmary to see her father and his colleagues. Kim, the camp’s nurse, informs her that their condition is not good. Her father seems to be much worse than the others. Kim is baffled by the illness and does not know what could have caused it. She asks Alyx to help her identify the disease and its source.
In this episode, the user takes control of her character upon arriving at the camp’s infirmary, which is housed in the same building as the laboratory. All of the user’s interactions, as she works with the camp nurse, occur within the confines of the infirmary and laboratory. A typical scenario has the user learning about the scientific method (Figure 4a), examining patients to learn about their symptoms (Figure 4b), learning about infectious diseases by reading books (Figure 4c), testing food items to find out which ones are contaminated (Figure 4d), convincing the camp nurse of a diagnosis, and finally treating the sickened research team members for their illness. Figure 4 depicts a plot graph (Bates 1992) of this Crystal Island episode.
Wizard-of-Oz Functionality

In the WOZ-enabled version of Crystal Island, a wizard provides the narrative planning functionalities, including spoken natural language dialogue for the character of the camp nurse. Playing the role of the camp nurse, the wizard works collaboratively with the user to solve the science mystery. Together in the virtual environment they carry on rich conversations using voice chat, and they observe one another’s actions while engaging in problem-solving activities (Figure 5). In addition to directing the navigation, spoken communication, and manipulation behaviors of the nurse’s character in the virtual environment, the wizard guides the user’s inquiry activities and controls the progression of
the story. To support these activities, the wizard’s display includes detailed information regarding the user’s activities in the environment (e.g., reading books, testing objects, updating the diagnosis worksheet) as well as access to a narrative dashboard. The narrative dashboard enables the wizard to initiate key narrative decisions in the environment (e.g., introducing new patient symptoms, having a non-player character bring in additional items for testing) in the same manner that a director agent acts.

In addition to the wizard functionalities, the narrative environment was modified to focus on the rich interactions between the user and wizard as well as to reduce the time spent navigating the environment. This was accomplished by confining the scenario to a single building housing both the camp’s infirmary and laboratory. Within this environment the user 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, user and wizard controls, and wizard display were refined throughout a series of pilot studies with college students prior to the corpus collection reported in this paper.

4.3 Example Scenario
To illustrate the behavior of the WOZ-enabled CRYSTAL ISLAND environment, consider the following scenario. A user has been collaborating with the nurse character, whose behaviors are planned and executed by the wizard. The user 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 user 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, “Yeah, hum, well, they both can come from water, but cholera is mostly water I believe,” the wizard determines that the user is having difficulty ruling out shigellosis and decides that this is an opportune moment to provide a hint. The wizard uses the narrative dashboard and enables the Introduce-Leg-Cramps symptom plot point, which results in one of the patients moaning loudly in the infirmary. The user examines the patient and informs the wizard, “He has leg cramps. That means it is cholera.” The wizard asks the user to update her diagnosis worksheet with her new hypothesis and explain why she believes this. The user then provides a detailed explanation justifying her diagnosis, and the story concludes with the nurse treating the patients for cholera.

4.4 Narrative Observations
Narrative is a complex phenomenon, and interactive narrative is becoming increasingly sophisticated. Director agent decision-making strategies should be carried out while considering different aspects of narrative states. In our investigation of decision-making strategies we consider multiple types of narrative observations. Below we examine the role of narrative observations and discuss the role that they can play in director agent decision-making strategies.
Table 1. Narrative action decisions

<table>
<thead>
<tr>
<th>Decisions</th>
<th>Tutorial Type</th>
<th>Descriptions</th>
<th>Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>START-SESSION</strong></td>
<td>Define Problem</td>
<td>Wizard gives a brief explanation of the student’s objectives and goals.</td>
<td>6.2%</td>
</tr>
<tr>
<td><strong>INTRODUCE-SCIENTIFIC-METHOD</strong></td>
<td>Background Information</td>
<td>Wizard explains to the student and suggests they use the scientific method while diagnosing the mysterious illness.</td>
<td>6.2%</td>
</tr>
<tr>
<td><strong>INTRODUCE-WORKSHEET</strong></td>
<td>Background Information</td>
<td>Wizard explains usage of the diagnosis worksheet to help the student formulate and refine their hypothesis.</td>
<td>6.2%</td>
</tr>
<tr>
<td><strong>EXAMINE-PATIENT-SYMPTOMS</strong></td>
<td>Hint</td>
<td>Wizard and student work together to examine symptoms of each of the patients.</td>
<td>8.1%</td>
</tr>
<tr>
<td><strong>UPDATE-WORKSHEET</strong></td>
<td>Confirm Understanding</td>
<td>Wizard reminds the student to update the diagnosis worksheet with new knowledge and hypothesis.</td>
<td>13.7%</td>
</tr>
<tr>
<td><strong>READ-DISEASE-BOOKS</strong></td>
<td>Hint/Advice</td>
<td>Wizard guides the student to read relevant disease information in the library, which helps them refine their hypothesis.</td>
<td>13.9%</td>
</tr>
<tr>
<td><strong>INTRODUCE-HEADACHE</strong></td>
<td>Hint</td>
<td>Wizard triggers an action resulting in a patient moaning and complaining about having a headache.</td>
<td>6.2%</td>
</tr>
<tr>
<td><strong>TEST-CAMP-ITEMS</strong></td>
<td>Advice</td>
<td>Student and wizard test food items the expedition team took with them from camp.</td>
<td>5.4%</td>
</tr>
<tr>
<td><strong>TEST-OUTSIDE-CAMP-ITEMS</strong></td>
<td>Advice</td>
<td>Student and wizard test food items the team found during their expedition.</td>
<td>3.4%</td>
</tr>
<tr>
<td><strong>TEST-CONTAMINATED-BANANAS</strong></td>
<td>Advice</td>
<td>Student and wizard test the bananas, which end up being contaminated.</td>
<td>3.4%</td>
</tr>
<tr>
<td><strong>INTRODUCE-DIRTY-WATER</strong></td>
<td>Hint</td>
<td>Wizard triggers an event causing a door to open and a water bottle to appear in the infirmary room.</td>
<td>5.2%</td>
</tr>
<tr>
<td><strong>INTRODUCE-LEG-CRAMPS</strong></td>
<td>Hint</td>
<td>Wizard triggers an event causing one of the patients to complain about leg cramps.</td>
<td>3.6%</td>
</tr>
<tr>
<td><strong>COMPLETE-WORKSHEET</strong></td>
<td>Confirm Understanding</td>
<td>Wizard asks student to update all remaining information that has not been entered and formulate their final hypothesis.</td>
<td>6.4%</td>
</tr>
<tr>
<td><strong>REPORT-RESOLUTION</strong></td>
<td>Confirm Understanding</td>
<td>Wizard asks student to explain their final hypothesis and how they arrived at their conclusion using the scientific method.</td>
<td>6.2%</td>
</tr>
<tr>
<td><strong>END-SESSION</strong></td>
<td>Confirm Understanding</td>
<td>Wizard thanks student and tells her that the patients will be treated based on her finding.</td>
<td>6.2%</td>
</tr>
</tbody>
</table>
4.4.1 Narrative Action Decisions
The wizard controls the progression of the story and scaffolds student interactions by utilizing the narrative dashboard. The narrative dashboard enables the wizard to initiate key narrative action decisions in the environment analogous to narrative planners. Table 1 describes the decisions that can be enacted by the wizard using the narrative dashboard.

There are fifteen narrative decisions that wizards can enact in the environment. Table 1 also summarizes the relative frequency of each decision, i.e., the ratio of the number of occurrences of specific decisions to the total number of decisions in all sessions. The frequencies range from 3.4% to 13.9% ($M = 6.7\%$, $SD = 3.2\%$). The corpus collection environment records detailed logs of actions performed by the student and wizard within the virtual environment, including decisions made by the wizards using the narrative dashboard. These logs provide a rich source of data to build empirically driven models of narrative decision-making.

4.4.2 Narrative Progress
The narrative arc framework (Figure 6) can be used to analyze narratives. A narrative arc models the tension experienced by the audience as a narrative progresses through its phases of exposition, complication, escalation, climax, and resolution. In the exposition phase, the setting and situation are introduced. During the complication phase, a problem develops and tension rises. The escalation phase sees the problem intensify and a rapid rise in the tension. The tension reaches its highest level during the climax phase when the story starts to resolve itself. During the resolution phase the remaining issues are resolved and the tension diminishes. Table 2 shows the narrative arc phases and their duration as defined within this Crystal Island episode. Utilizing the current phase of the narrative arc as an observation during narrative decision-making provides evidence regarding the high level structure of the unfolding narrative.
Table 2. Narrative arc phases

<table>
<thead>
<tr>
<th>Narrative Arc Phases</th>
<th>Duration of Narrative Arc Phases in CRYSTAL ISLAND</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposition</td>
<td>From Start-Session to beginning of the first occurrence of Examine-Patient-Symptoms</td>
</tr>
<tr>
<td>Complication</td>
<td>From the end of Exposition to beginning of the first occurrence of Read-Disease-Books</td>
</tr>
<tr>
<td>Escalation</td>
<td>From the end of Complication to the first occurrence of Introduce-Headache or Test-Contaminated-Bananas (whichever comes first)</td>
</tr>
<tr>
<td>Climax</td>
<td>From the end of Escalation to the first occurrence of Introduce-Dirty-Water or Introduce-Leg-Cramp (whichever comes latter)</td>
</tr>
<tr>
<td>Resolution</td>
<td>From the end of Climax to End-Session</td>
</tr>
</tbody>
</table>

4.4.3 Physical State
The physical state indicates the location of characters in the storyworld (i.e., student, wizard), which are represented as discretized virtual world locations. Information about the location of the user and wizard within the storyworld also plays an important role in narrative decision-making. For example, narrative decisions such as Introduce-Scientific-Method often occur in the location where a poster about the scientific method is hanging on
the wall. In this episode, all of the user’s interactions, as she works with the camp nurse, occur within the infirmary and laboratory, which are housed in the same building. Table 3 presents a detailed description of the locations within the building. Utilizing the current location of the user and/or wizard as an observation during narrative decision-making provides details regarding likely directions for the narrative.

<table>
<thead>
<tr>
<th>Locations</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infirmary-Poster</td>
<td>Place where the scientific method poster is located. Wizard uses the poster to elucidate the detail of scientific method to the user.</td>
</tr>
<tr>
<td>Infirmary-Bryce-Bed</td>
<td>Place where patient Bryce is located.</td>
</tr>
<tr>
<td>Infirmary-Ford-Bed</td>
<td>Place where patient Ford is located.</td>
</tr>
<tr>
<td>Infirmary-Audrey-Bed</td>
<td>Place where patient Audrey is located.</td>
</tr>
<tr>
<td>Infirmary-Desk</td>
<td>Place where cook sets dirty water bottle.</td>
</tr>
<tr>
<td>Lab-Table-A</td>
<td>Place where camp food items are located and ready to test.</td>
</tr>
<tr>
<td>Lab-Table-B</td>
<td>Place where outside camp food items are located and ready to test.</td>
</tr>
<tr>
<td>Lab-Testing-Equipment</td>
<td>Place where food items can be tested to see if it is infected.</td>
</tr>
<tr>
<td>Lab-Library</td>
<td>Place where relevant disease books are located.</td>
</tr>
</tbody>
</table>

4.5 Corpus Collection Procedure

In the data collection, more than twenty hours of trace data were collected using the WOZ-enabled CRYSTAL ISLAND environment. The trace data includes detailed logs of all the user and wizard actions (e.g., navigation, manipulation, and decision making) within the environment, as well as audio and video recordings of their conversation.

4.5.1 Participants

The participants were 33 eighth-grade students (15 males and 18 females) from a public school in 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.

4.5.2 Participant Procedure
When users 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 Crystal Island back-story, task description, characters, and controls. Upon completing their review of the handouts, the researcher provided further direction to the users on the use of the keyboard and mouse controls (see Appendix A for additional details). The researcher then informed the users that they would be collaborating with another human-controlled character, the camp nurse, in the environment to solve the science mystery. Users were asked to communicate with the camp nurse throughout their sessions. Finally, the researcher answered any questions from the users, 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 user for the duration of their session. The Crystal Island session concluded once the user and wizard arrived at a treatment plan for the sickened researcher. The users’ sessions lasted no more than sixty minutes ($M = 38, SD = 5.15$). During model evaluation one of the participants was eliminated as an outlier—the data were more than three standard deviations from the mean—leaving thirty-two usable trace data logs.

4.5.3 Wizard Protocol
To improve the consistency of the wizards’ tutorial planning, narrative planning, and natural language dialogue activities, a protocol was iteratively developed and refined through a series of pilot studies (see Appendix B for additional details). The resulting protocol included a high-level procedure for the wizard to follow (e.g., introduce yourself as the camp nurse, describe the patient situation to the student, review the scientific method with the student), a set of interaction guidelines (e.g., collaboratively work with the student to solve the mystery,
organize the student’s activities around the scientific method, act as a senior peer to the student, encourage the student to explain her conclusions and ensure they are logical and consistent with the available data, engage the student in constant face-to-face inquiry dialogue), and a set of narrative guidelines (e.g., overall story structure, appropriate contexts for narrative decisions, ordering constrains, be cognizant of the elapsed time to ensure the session completes in a timely manner).

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.
Chapter 5

Modeling Director Agents with Dynamic Bayesian Networks

Utilizing the resulting corpus of trace data from the corpus collection we can learn models of director agents decision-making strategies for both narrative intervention and action using the proposed dynamic Bayesian network structure. Director agents utilize numerous storyworld observations that change over time to accurately determine the most appropriate time to intervene and the next director agent actions in the unfolding story. A DBN is a directed acyclic graph that incorporates time slices, where each time slice contains its own state variables. By utilizing time slices, DBNs support probabilistic inference for events that change over time.

The models were implemented with the GeNle/SMILE Bayesian modeling and inference library developed at the University of Pittsburgh’s Decision System Laboratory (Druzdzel 1999). Given the network structure of the DBN, the probabilities of each node in the network were learned by performing parameter learning for the conditional probability tables (CPTs). The Expectation-Maximization algorithm from the SMILearn library was used to learn the CPT parameters. After CPT parameters were learned, the resulting network was used to
make inferences about the director agent action and intervention decision nodes in the models.

It should be noted that the particular conditional probabilities learned as the parameters for the models are specific to the story arc, characters, and virtual environment of CRYSTAL ISLAND. However, it appears that the methodology is generalizable across narrative spaces and virtual environments. To learn the parameters for a new model, a model can be trained using the Wizard-of-Oz methodology for a new narrative space and environment, the resulting model can be used to make intervention and action decision-making strategies for that narrative space and virtual environment.

5.1 Feature Selection
Feature selection problem has been studied many years by the machine learning community (Kira & Rendell 1992; Koller & Sahami 1996; Kohavi & John 1997; Jebara & Jaakkola 2000). By selecting the most relevant features from the data, the performance of learning models can be significantly improved. For our study we utilized a two-step feature selection approach to devise accurate models of director agents’ intervention and action decision strategies. First, we performed a post-study “after action review” with the wizard to discuss his or her narrative decision strategies while interacting with the students. During the corpus collection we video recorded all of the actions taken by the wizards within the WOZ-enabled CRYSTAL ISLAND. After the corpus collection we asked the wizards to watch the recorded videos and describe why the narrative decisions were chosen and what features were accounted for when they were making the narrative decisions. Based on input from the wizards’ post-study “after action review,” we selected an initial set of features for the intervention and action decision strategies models. Second, we performed a brute force feature ranking method. Selected features were evaluated with all possible combinations of the input features. The subsets with the features that yielded the most efficient performance
for each intervention and action decision strategies models were chosen and implemented in the models.

5.2 Intervention Decision Strategies

The high-level structure of the dynamic Bayesian network model created for director agent intervention decision strategies is shown in Figure 7. Three time-slices are illustrated in the figure with the intervention decision from the previous time slice, $\text{intervention decision}_{t-1}$, influencing the current intervention decision, $\text{intervention decision}_t$. Within each time slice, observations from the story world, $\text{narrative state}_t$, also influence the intervention decision. These observations include items such as 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 belief about the overall state of the narrative.

The DBN model consists of the following elements:

- **Intervention Decision**: Models the director agent intervention decision. The intervention decision is a binary variable taking on the values of either action or
no-action. Action indicates that a director action should be taken to intervene in the story while no-action indicates that the director agent should remain inactive. In the network, the beliefs about intervention decision

\( t-1 \) in time slice \( t-1 \) are influenced by physical state

\( t-1 \), narrative progress

\( t-1 \), user knowledge

\( t-1 \), and time index

\( t-1 \). The intervention decision

\( t-1 \) influences physical state

\( t \), narrative progress

\( t \), and user knowledge

\( t \) which in-turn influence intervention decision

\( t \) in time slice \( t \).

- **Physical State**: Models the current location of the user and wizard in the storyworld’s virtual environment. In the network, the current beliefs about physical state

\( t \) in time slice \( t \) are influenced by the physical state

\( t-1 \) and intervention decision

\( t-1 \) in time slice \( t-1 \) and influences the intervention decision

\( t \) in time slice \( t \).

- **Narrative Progress**: Models the storyworld’s narrative structure. To characterize the progress of the narrative, we analyzed the story structure utilizing a narrative arc framework. Utilizing the current phase of the narrative arc as an observation provides the model with evidence about the high level structure of the unfolding narrative. In the network, the current beliefs about narrative progress

\( t \) in time slice \( t \) are influenced by the narrative progress

\( t-1 \) and intervention decision

\( t-1 \) in time slice \( t-1 \) and influences the intervention decision

\( t \) in time slice \( t \).

- **User Knowledge**: Models the user’s beliefs about the salient facts of the story learned through interactions with the environment and other characters. Within the CRYSTAL ISLAND environment users complete a diagnosis worksheet while solving the science mystery (Shores, Rowe, & Lester 2011), which provides details regarding users’ current beliefs about the story. In the network, the current beliefs about user knowledge

\( t \) in time slice \( t \) are influenced by the user knowledge

\( t-1 \) and intervention decision

\( t-1 \) in time slice \( t-1 \) and influences the intervention decision

\( t \).

- **Time Index**: Models the overall timeline of the storyworld to provide the temporal-based evidence for the intervention decision. In the network, time index

\( t \) influences
the current intervention decision. It is not influenced by other observable variables since it is a deterministic monotonically increasing sequence.

At runtime, as new observations become available, such as user and wizard locations, the corresponding nodes in the network are updated with their observed values. Influences are then propagated throughout the network, allowing inferences to be made regarding the most probable intervention decision at time slice $t$.

5.3 Action Decision Strategies
The high-level structure of the dynamic Bayesian network model created for narrative decision-making is shown in Figure 8. The figure illustrates three time slices and their corresponding narrative action decisions: $action\ decision_{t-2}$, $action\ decision_{t-1}$, and $action\ decision_t$. The three time slices include representations of the narrative observation including information on the physical state of the storyworld and progression of the narrative. Each time slice encodes a probabilistic representation of the belief about the overall state of the narrative.

The action decision nodes model the knowledge of prior action decisions. The physical state nodes model the location of characters in the storyworld (i.e., student, wizard), which are represented as discretized virtual world locations. The narrative progress nodes model the storyworld’s narrative structure. To characterize the progress of the narrative, we analyzed the story structure utilizing a narrative arc framework. Utilizing the current phase of the narrative arc as an observation provides the model with evidence about the high level structure of the unfolding narrative. The model considers the current beliefs about the physical state and narrative progress represented in narrative observation. It also considers prior history of $action\ decision_{t-1}$ and $action\ decision_{t-2}$. Using the links from $action\ decision_{t-1}$, $action\ decision_{t-2}$, narrative state, and narrative progress, the model captures how each of these influences action decision.
Given the DBN structure, the values in the conditional probability tables (CPTs) for each observation node in the network can be learned using a corpus. Setting observed evidence on the learned model and updating the network allows the likelihood of decisions to be computed at each time slice.

### 5.4 Integrated Real-Time Models

To explore the real-time effectiveness of the machine-learned models of the director agent decision-making strategies, the intervention model and the action model were integrated into the CRYSTAL ISLAND story-based learning environment (Figure 9). The environment is identical to the WOZ-enabled CRYSTAL ISLAND except non-player character interactions are driven by the director agent (Figure 10). Users interact with non-player characters to receive environmental information (e.g., How do you operate the testing equipment? or Where is the library?), and microbiology concepts (e.g., What is a waterborne disease?) using multimodal dialogue. Users select their questions using a dialogue menu and characters respond with spoken language.
Actions generated by the director agent decision models are primarily initiated via the camp nurse. For example, when the model determines that it is an appropriate time to intervene to help the user examine patient symptoms to solve the mystery, the model directs the camp nurse to walk to the user. The camp nurse, using spoken language, informs the user that she should examine patients to determine their current symptoms. The camp nurse then guides the user to the infirmary to examine the patients.

To solve the mystery, students complete a *diagnosis worksheet* to organize their hypotheses and record findings about patient symptoms and testing results. Once users have completed their diagnosis worksheets with the source and cause of the illness, they can submit their solutions to the camp nurse for review.
To illustrate the behavior of the machine-learned models of the director agent decision-making strategies, consider the following scenario. A user has been interacting with the nurse character (a), whose behaviors are planned and executed by the director agent decision-making strategies models. Under the guidance of the camp nurse, the user has examined the patients’ symptoms (b) and read reference books about diseases (c). Through this exploration users were able to narrow down their possible disease hypotheses. Currently, the user is in the library reading disease reference books (d). With observations of current user location, what she has achieved, how much she has learned via the diagnosis worksheet (e), and the overall story progression, the director agent intervention decision strategies model determines it is the opportune moment to provide a hint. The director agent action decision model enables the patient moaning sound and directs the camp nurse to explain that she just heard the patient moaning and that the user should go and check the patients (f). The camp nurse guides the user to the infirmary (g) and examines the patient for their new symptoms.
Figure 12. CRYSTAL ISLAND real-time interaction
Chapter 6

Evaluation

To explore the effectiveness of our models of director agent decision-making strategies, we have conducted an empirical evaluation. Although intelligent narrative systems offer significant promise for interactive narrative environments, designing the director agent decision-making strategies presents serious computational challenges to achieve effective narrative experiences for users.

6.1 Hypotheses

With a focus on creating engaging narrative experiences and pedagogically effective story-based learning for users in narrative-centered learning environments, the following hypotheses were tested in an evaluation study to assess the pedagogical effectiveness of the learned director agent decision-making models as well as the models effect on engagement and in-game performance:
Hypothesis I: Empirically informed director agent decision-making models operating within a narrative-centered learning environment with full guidance can achieve higher user learning gains than intermediate and minimal director agent guidance.

Hypothesis II: Empirically informed director agent decision-making models operating within a narrative-centered learning environment with full guidance can achieve higher in-game performance than intermediate and minimal director agent guidance.

Hypothesis III: Empirically informed director agent decision-making models operating within a narrative-centered learning environment with full guidance can achieve higher user engagement and interest than intermediate and minimal director agent guidance.

Hypothesis IV: Empirically informed director agent decision-making models for a narrative-centered learning environment can reduce development efforts in comparison with similar human-authored models.

6.2 Predictive Accuracy Evaluation
The Performance of the empirically driven models of the director agent intervention and action decisions was evaluated with respect to predictive accuracy. Results were encouraging and suggested that empirically driven models of director agent decision-making strategies can offer significant predictive power.

6.2.1 Intervention Decision Model
For the DBN model, there are a total of 84 time slices, 420 nodes, and more than 5200 conditional probabilities present in the director agent intervention decision-making network. The number of slices was determined based on the smallest time interval between narrative interventions found in the collected corpus.
6.2.1.1 Results and Discussion

An analysis was conducted to assess using dynamic Bayesian networks for modeling director agent intervention strategies. To compare the effectiveness of the DBN model, a naïve Bayes model was developed as a baseline in which all observable variables are assumed to be independent of one another. Both of the models were learned using trace data collected from thirty-two interactive CRYSTAL ISLAND sessions. A leave-one-out cross validation method was employed to ensure the learned models were not over-fitted. Recall, precision, and accuracy were computed using an aggregated confusion matrix for each model.

<table>
<thead>
<tr>
<th>Narrative Intervention Model</th>
<th>Recall</th>
<th>Precision</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naïve Bayes</td>
<td>39.3%</td>
<td>31.9%</td>
<td>75.5%</td>
</tr>
<tr>
<td>DBN</td>
<td>73.3%</td>
<td>82.0%</td>
<td>92.8%</td>
</tr>
</tbody>
</table>

Table 4 shows the results of classification measurements for the naïve Bayes and DBN models. It was found that the DBN model outperformed the baseline model naïve Bayes model in all categories. There are significant improvements exhibited by the DBN model. The DBN model achieved a prediction accuracy of 92.8% and the baseline achieved 75.5%. The DBN model exhibited a more than 16% accuracy improvement over the baseline. Also, the DBN model achieves significant gains on both recall and precision, 34% and 50% respectively, as compared to the baseline.

The evaluation indicates that using empirically informed dynamic Bayesian network models for director agent intervention strategies is a promising approach. It was found that the DBN model significantly outperformed the baseline model in all classification analysis. The results suggest that in interactive narrative environments the independence assumption underlying naïve Bayes models may not hold: it appears that providing evidence regarding narrative structure, physical locations, the user’s beliefs, overall story timeline, intervention decision
history, and their dependent relationships can significantly improve director agent intervention decision predictions.

### 6.2.2 Action Decision Model

For the DBN model, there are a total of 22 time slices, 88 nodes, and more than 830 conditional probabilities present in the narrative decision-making network.

#### 6.2.2.1 Results and Discussion

An analysis was conducted to investigate the use of dynamic Bayesian networks for modeling narrative decision-making. To compare the effectiveness of the DBN model against a baseline, a bi-gram model was developed in which only the previous action decision was used to predict the next action decision. The results are presented in terms of overall accuracy, macro-averaged recall, and macro-averaged precision as shown in Table 5. Since the action decision model solves a multi-class classification problem, we computed the results using macro-averaged recall and precision (Yang & Liu 1999; Özgür, Özgür, & Güngör 2005). Macro-averaging is computed by taking the arithmetic means of all classes and dividing the sum of the recall or precision over all classes by the total number of classes presents results of the macro-averaged recall or precision. Macro-averaging assumes each class has equal weight. This performance metric is commonly used in natural language processing.

<table>
<thead>
<tr>
<th>Narrative Action Model</th>
<th>Macro-averaged Recall</th>
<th>Macro-averaged Precision</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>bi-gram</td>
<td>74.0%</td>
<td>73.0%</td>
<td>71.0%</td>
</tr>
<tr>
<td>DBN</td>
<td>93.0%</td>
<td>94.0%</td>
<td>93.7%</td>
</tr>
</tbody>
</table>

The bi-gram model achieved an action decision predictive accuracy of 71%. A leave-one-out cross validation method was employed. In the prediction evaluation, the DBN model achieves action decision prediction accuracy of 93.7%. The DBN model exhibited a 23% accuracy improvement over the bi-gram model. Also, the DBN model achieves significant
gains on both recall and precision by 20% as compared to the baseline. It appears that providing evidence regarding narrative structure, physical locations, and action decision history can significantly improve action decision prediction.

6.3 Runtime Evaluation
To investigate the runtime effectiveness of our models of director agent decision-making strategies, we have conducted an empirical evaluation. With a focus on creating engaging narrative experiences and pedagogically effective story-based learning for users in interactive narrative-centered learning environments, the following research design was proposed.

6.3.1 Study Design Considerations
Several evaluation schemes were considered. These included items such as using wizard-based interaction instead of a fully automated system, investigating authoring efficiency, exploring ablated models, as well as comparing against random and fixed policies.

A key consideration for the evaluation is whether to use a fully automated runtime system or to use a wizard-based system (i.e., an evaluation in which a wizard controls the nurse character’s navigation and communication behaviors but the director agent controls the narrative progression). Several potential advantages of a wizard-based approach were explored. First, a wizard-based approach would require less software development effort since the AI to control the nurse character would be unnecessary. Second, the wizard-based approach would more closely mirror the system used during the data collection study which might allow comparisons to be made. However, even with the potential advantages a wizard-based approach for the runtime study was determined to be infeasible for several reasons. First, it would be very challenging to keep wizards from using or relying on their previous session experiences, which would potentially contaminate future sessions and could cause undesirable interactions between conditions (i.e., the wizard might accidently use knowledge from one condition while interacting in another condition since wizards would likely need to randomly participate in all conditions). Second, inconsistent wizard actions and dialogue
could impact the results of the study (e.g., dialogue and behaviors of wizard would vary one session to another due to fatigue and other external influences). Third, it is questionable that any meaningful comparisons between the original Wizard-of-Oz data collection could be made with results from the new study since interaction between wizard and user would be different due to the wizard relinquishing the control of narrative to the learned model. Finally, and perhaps most critically, a wizard-based approach for the runtime evaluation would be infeasible. A wizard-based approach would require a wizard to be available for each student interaction which greatly reduces throughput (i.e., only a few sessions could be run at a time instead of up-to 60 students at a time). Given that the evaluation will likely involve on the order of 120 participants, a wizard-based approach is not possible (e.g., working the study into the public school schedule, finding rooms at a school for the duration of the study, finding enough wizard time to conduct the study).

Another consideration for the evaluation is whether to use a fixed policy (such as a finite state machine) or random policy. A fixed policy was determined to be undesirable. It is difficult to verify the quality of using a fixed policy since it is unlikely that the results of utilizing the fixed policy would produce similar performance when applied in other settings. Utilizing a random policy was also discarded since it has the potential to generate an incoherent narrative and would offer a comparison case that is “too low.”

Another consideration for the evaluation was whether to use an ablation study. An ablation study that considers different levels of director agent guidance was considered. Similar experimental designs have been used in recent game-based learning environment (Thomas, 2011). In Thomas’ dissertation study, he considers different levels of assistance from the system to help students to achieve their goals. This approach appears to be promising for determining the effectiveness of the director agent decision-making models.
6.3.2 Evaluation Design

To explore the first three principal hypotheses noted above, an empirical approach to evaluation was performed. The following research design was used to investigate the hypotheses. We will employ three different conditions: Minimal Guidance, Intermediate Guidance, and Full Guidance.

- **Minimal Guidance**: Subjects experience the storyworld controlled by a minimal director agent decision-making model. This model includes the actions that must be achieved by the director agent (i.e., the user cannot achieve them without the system taking action). The model in this condition was not learned; rather, it makes decisions once all preconditions are met for an action to be taken. The minimal director agent decision-making model employs the five director action decisions listed in Table 6.

<table>
<thead>
<tr>
<th>Decisions</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>START-SESSION</td>
<td>Wizard gives a brief explanation of the student’s objectives and goals.</td>
</tr>
<tr>
<td>INTRODUCE-HEADACHE</td>
<td>Wizard triggers an action resulting in a patient moaning and complaining about having a headache.</td>
</tr>
<tr>
<td>INTRODUCE-DIRTY-WATER</td>
<td>Wizard triggers an event causing a door to open and a water bottle to appear in the infirmary room.</td>
</tr>
<tr>
<td>INTRODUCE-LEG-CRAMPS</td>
<td>Wizard triggers an event causing one of the patients to complain about leg cramps.</td>
</tr>
<tr>
<td>END-SESSION</td>
<td>Wizard thanks student and tells her that the patients will be treated based on her finding.</td>
</tr>
</tbody>
</table>

- **Intermediate Guidance**: Subjects experience the storyworld controlled by an intermediate director agent decision-making model. This is an ablated model that is inspired by the notions of *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 preconditions, the plan will never achieve its goal. Islands must occur at some intermediate time for
achieving the overall goals. In this version of CRYSTAL ISLAND, the transitions between narrative arc phases represent “islands” in the narrative. Table 7 outlines the director agent actions that transition between the narrative arc phases. Each arc phase consists of a number of potential director action decisions; however, the phases are bounded by specific director action decisions that define when each phase starts and ends. We employ these specific director action decisions as our islands. The intermediate director agent decision-making model employ only the eight director action decisions listed in Table 7.

Table 7. Set of director action decisions as islands

<table>
<thead>
<tr>
<th>Decisions</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>START-SESSION</td>
<td>Beginning of Exposition narrative arc phase in CRYSTAL ISLAND</td>
</tr>
<tr>
<td>EXAMINE-PATIENT-SYMPTOMS</td>
<td>End of Exposition phase and beginning of Complication narrative arc phase in CRYSTAL ISLAND</td>
</tr>
<tr>
<td>READ-DISEASE-BOOKS</td>
<td>End of Complication phase and beginning of Escalation narrative arc phase in CRYSTAL ISLAND</td>
</tr>
<tr>
<td>INTRODUCE-HEADACHE or TEST-CONTAMINATED-BANANAS</td>
<td>End of Escalation phase and beginning of Climax narrative arc phase in CRYSTAL ISLAND (whichever comes first)</td>
</tr>
<tr>
<td>INTRODUCE-DIRTY-WATER or INTRODUCE-LEG-CRAMPS</td>
<td>End of Climax phase and beginning of Resolution narrative arc phase in CRYSTAL ISLAND (whichever comes latter)</td>
</tr>
<tr>
<td>END-SESSION</td>
<td>End of Resolution phase in CRYSTAL ISLAND</td>
</tr>
</tbody>
</table>

- **Full Guidance**: Subjects experience the storyworld controlled by the full director agent decision-making model. The director agent actively monitors subjects interacting with the storyworld to determine when it is appropriate to intervene with the next director agent action to guide subjects to the intended narrative. The director agent has full control of the director intervention decisions (i.e., “when to intervene”) and director action
decisions (i.e., “how to intervene”). The full guidance director agent decision-making model will employ all of the director action decisions listed in Table 8.

<table>
<thead>
<tr>
<th>Decisions</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>START-SESSION</strong></td>
<td>Wizard gives a brief explanation of the student’s objectives and goals.</td>
</tr>
<tr>
<td><strong>INTRODUCE-SCIENTIFIC-METHOD</strong></td>
<td>Wizard explains to the student and suggests they use the scientific method while diagnosing the mysterious illness.</td>
</tr>
<tr>
<td><strong>INTRODUCE-WORKSHEET</strong></td>
<td>Wizard explains usage of the diagnosis worksheet to help the student formulate and refine their hypothesis.</td>
</tr>
<tr>
<td><strong>EXAMINE-PATIENT-SYMPTOMS</strong></td>
<td>Wizard and student work together to examine symptoms of each of the patients.</td>
</tr>
<tr>
<td><strong>UPDATE-WORKSHEET</strong></td>
<td>Wizard reminds the student to update the diagnosis worksheet with new knowledge and hypothesis.</td>
</tr>
<tr>
<td><strong>READ-DISEASE-BOOKS</strong></td>
<td>Wizard guides the student to read relevant disease information in the library, which helps them refine their hypothesis.</td>
</tr>
<tr>
<td><strong>INTRODUCE-HEADACHE</strong></td>
<td>Wizard triggers an action resulting in a patient moaning and complaining about having a headache.</td>
</tr>
<tr>
<td><strong>TEST-CAMP-ITEMS</strong></td>
<td>Student and wizard test food items the expedition team took with them from camp.</td>
</tr>
<tr>
<td><strong>TEST-OUTSIDE-CAMP-ITEMS</strong></td>
<td>Student and wizard test food items the team found during their expedition.</td>
</tr>
<tr>
<td><strong>TEST-CONTAMINATED-BANANAS</strong></td>
<td>Student and wizard test the bananas, which end up being contaminated.</td>
</tr>
<tr>
<td><strong>INTRODUCE-DIRTY-WATER</strong></td>
<td>Wizard triggers an event causing a door to open and a water bottle to appear in the infirmary room.</td>
</tr>
<tr>
<td><strong>INTRODUCE-LEG-CRAMPS</strong></td>
<td>Wizard triggers an event causing one of the patients to complain about leg cramps.</td>
</tr>
<tr>
<td><strong>COMPLETE-WORKSHEET</strong></td>
<td>Wizard asks student to update all remaining information that has not been entered and formulate their final hypothesis.</td>
</tr>
<tr>
<td><strong>REPORT-RESOLUTION</strong></td>
<td>Wizard asks student to explain their final hypothesis and how they arrived at their conclusion using the scientific method.</td>
</tr>
<tr>
<td><strong>END-SESSION</strong></td>
<td>Wizard thanks student and tells her that the patients will be treated based on her finding.</td>
</tr>
</tbody>
</table>
6.3.3 Method
In this section we present a description of the participants and their participation procedure. The material and apparatus that were used for the study to evaluate the hypotheses are also discussed.

6.3.3.1 Participants
A total of 183 students interacted with CRYSTAL ISLAND. Participants were all eighth-grade students from North Carolina State public school ranging in age from 12 to 15 ($M = 13.40$, $SD = 0.53$). Twelve of the participants were eliminated due to hardware and software issues and twenty-one participants were eliminated due to incomplete data on either their pre-test or post-test. Among the remaining students, 68 were male and 82 were female. Approximately 1% of the students were American Indian or Alaska Native, 6% were Asian, 35% were Black or African American, 11% were Hispanic or Latino, 41% were Caucasian, and 6% were of other races. Table 9 shows the demographics of each condition.

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Gender</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>Full</td>
<td>28</td>
<td>27</td>
</tr>
<tr>
<td>Intermediate</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>Minimal</td>
<td>17</td>
<td>30</td>
</tr>
</tbody>
</table>

6.3.3.2 Materials and Apparatus
To evaluate Hypothesis 1, we used a content test that focuses on CRYSTAL ISLAND microbiology curriculum and the scientific method. Learning gains will be measured by the difference of post-test and pre-test scores.

To evaluate Hypothesis 2, we utilized several metrics to assess in-game performance: (1) number of subjects who complete the game; (2) efficiency of subjects achieving the goal; and (3) overall interaction time.
To evaluate Hypothesis 3, a variation of the Presence Questionnaire (Witmer & Jerome 2005) and the Subject Impressions Questionnaire from the Intrinsic Motivation Inventory (IMI) (Ryan, Mims, & Koestner 1983) were used. Presence describes a subject’s sense of being there when interacting in mediated environment (Witmer & Singer 1998). The Presence Questionnaire is a validated measure containing subscales, including subjects’ involvement, immersion, and interface quality. All measures consist of Likert scale questions. The involvement subscale is intended to assess the degree of attention and meaning that the individual attaches to some set of stimuli, activities, or events. Immersion subscale indicates subject’s continuous stream of sensory stimuli and experiences. Example items include the following: “Were you able to anticipate what would happen next in response to the actions that you performed?”, “How completely were your senses engaged in this experience?”, and “Were there moments during the virtual environment experience when you felt completely focused on the task or environment?” The interface quality measures the consistency and seamlessness of control and display devices that are incorporated with interactive environment. For example, items are “How much did your experiences in the virtual environment seem consistent with your real-world experiences?” and “How much did the visual display quality interfere or distract you from performing assigned tasks or required activities?” (Witmer & Singer 1998; Witmer & Jerome 2005).

The Intrinsic Motivation Inventory (IMI) is an instrument that assesses participating users’ subjective experiences while interacting with a mediated environment. The Subject Impressions Questionnaire from IMI is used to describe thoughts and feelings subjects may have regarding the director agent controlled character. The Subject Impression Questionnaire instrument assesses number of subscales, including users’ interest/enjoyment, perceived choice, effort, and pressure/tension (Ryan, Mims, & Koestner 1983). All measures consist of Likert scale questions. Subscales questionnaires are considered to be the self-report measure of intrinsic measure. Example items include Interest/Enjoyment (“While I was interacting with this person, I was thinking about how much I enjoyed it.”), Perceived Choice (“I felt
like I had choice about interacting with this person.”), Effort (“I tried hard to have a good interaction with this person.”), and Pressure/Tension (“I was anxious while interacting with this person.”)

For Hypothesis 4, we evaluated the development time efficiency of the data-driven director agent approach against a hand-authoring approach. A qualitative exploration was conducted to investigate the development time differences between these two approaches. Specifically, to compare these two approaches of director agent model building, we hand-authored parameters of the dynamic Bayesian networks (DBN) for director agent decision strategies and compared against the data-driven approach. Because manually authoring the entire DBN models would be impractical, we authored a slice for both the director intervention model and the director action model. There are a total of 84 slices in the director intervention model and 22 slices in the director action model. The slice to be hand-authored was chosen randomly. The resulting authoring time for a slice was multiplied by the total number of slices in each of the DBN models. The authoring time was compared against the overall time to build the empirically driven director agent models to analyze the development time efficiency.
6.3.3.3 Participant Procedure

Users entered the study room having completed pre-test materials several days prior to the intervention. Upon arriving, users were randomly assigned to a condition. Users were greeted by a researcher and provided details about CRYSTAL ISLAND during an introductory presentation by a member of the research team. After the presentation, users were instructed to review a set of CRYSTAL ISLAND instructional handouts, including information on the CRYSTAL ISLAND back-story, task description, characters, and controls. Upon completing their review of the handouts, the member of the research team provided further direction to the users on the use of the keyboard and mouse controls. Students were given 45 minutes to work on CRYSTAL ISLAND’s science mystery. Immediately after solving the mystery, or 45 minutes of interaction, students exited the CRYSTAL ISLAND learning environment and completed the post-test, which consisted of the same items as the pre-test. The post-test and
post materials were completed by the students within 30 minutes. In total, the students’ sessions lasted no more than 90 minutes. The study procedure is depicted in Figure 12.

6.3.4 Results
Analyses were conducted to assess the effectiveness of the director agent decision-making models by evaluating given hypotheses.

6.3.4.1 Hypothesis I: Learning Outcome and Differences
An investigation of overall learning found that students’ CRYSTAL ISLAND interactions yielded positive learning outcomes. A matched pairs t-test between post-test and pre-test scores indicates that the learning gains were significant, \( t(149) = 2.03, p < .05 \). Examining the learning outcome for each condition it was found that users’ CRYSTAL ISLAND interactions in the Full Guidance condition yielded significant learning gains, as measured by the difference of post-test and pre-test scores. A matched pairs t-test revealed that users in the Full Guidance condition showed statistically significant learning gains. Users in the Intermediate and Minimal Guidance conditions did not show significant learning gains (Table 10).

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Gain Avg</th>
<th>SD</th>
<th>( t )</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full</td>
<td>1.28</td>
<td>2.66</td>
<td>2.03</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>Intermediate</td>
<td>0.13</td>
<td>2.69</td>
<td>0.19</td>
<td>0.84</td>
</tr>
<tr>
<td>Minimal</td>
<td>0.89</td>
<td>3.12</td>
<td>1.23</td>
<td>0.22</td>
</tr>
</tbody>
</table>

In addition, there was a significant difference between the conditions in terms of learning gains. Controlling for pre-test scores using ANCOVA, the learning gains for the Full and Minimal Guidance conditions were significantly different, \( F(2, 99) = 38.64, p < .001 \) and the Full and Intermediate Guidance conditions were also significantly different, \( F(2, 100) = 40.22, p < .001 \). Thus, users in the Full Guidance condition achieved significantly higher learning gains than the users in the other two conditions.
6.3.4.2 Hypothesis II: In-Game Problem-Solving Effectiveness

To more closely investigate the effectiveness of the director agent decision-making strategies model, additional analyses for users’ in-game problem-solving performance were conducted. In order to compare the behavior of users problem-solving performances among the conditions, we investigated the users’ gameplay efficiency by analyzing whether they solved CRYSTAL ISLAND’s science mystery and their game completion time. Table 11 reports the game play performance for each condition.

Table 11. In-game problem-solving performances

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Solved Mystery</th>
<th>Completion Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>Full</td>
<td>92.73 %</td>
<td>1724</td>
</tr>
<tr>
<td>Intermediate</td>
<td>85.42 %</td>
<td>1761</td>
</tr>
<tr>
<td>Minimal</td>
<td>70.21 %</td>
<td>2229</td>
</tr>
</tbody>
</table>

To analyze the difference in the number of users who solved the mystery among the conditions, a chi-square test was performed. The results showed that the correlation is significant, (likelihood ratio, $\chi^2 = 9.37$, Pearson, $\chi^2 = 9.47$, $p < .01$), indicating that the number of users who solved the mystery varied significantly among the conditions. We also examined the differences in time it took students to solve the mystery. An ANOVA test was performed to investigate the differences among the conditions. The test revealed that differences were significant, $F(2, 122) = 15.13$, $p < .001$, which implied that the total time it took to solve the mystery varied significantly among the different conditions. Tukey’s pairwise comparison tests further indicated that the Full and Minimal Guidance conditions are significantly different ($p < .001$), as well as the Intermediate and Minimal Guidance conditions ($p < .001$). However, Tukey’s test did not reveal any significant differences between the Full and Intermediate Guidance conditions.

6.3.4.3 Hypothesis III: User Engagement and Interest

Analyses were also conducted to determine whether the participating users experienced different levels of engagement and interest while interacting with the CRYSTAL ISLAND
learning environment. Table 12 displays raw scores of presence measures for each condition. An ANOVA was performed to examine the differences among the conditions. The test revealed that difference was not significant in PQ’s involvement subscale, $F(2, 142) = 0.07, p = .930$. The PQ’s immersion and interface quality subscales were observed, $F(2, 142) = 1.16, p = .317; F(2, 142) = 1.24, p = .290$, respectively. No significant differences were found on all PQ’s subscale measures when compared on each condition.

Table 12. Means and SD on PQ raw scores

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Overall Presence</th>
<th>Presence Subscales</th>
<th>Interface Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Involvement</td>
<td>Immersion</td>
</tr>
<tr>
<td>Full</td>
<td>132.7</td>
<td>52.8 (12.9)</td>
<td>39.1 (7.87)</td>
</tr>
<tr>
<td>Intermediate</td>
<td>132.1</td>
<td>53.4 (14.8)</td>
<td>38.2 (9.97)</td>
</tr>
<tr>
<td>Minimal</td>
<td>129.0</td>
<td>52.3 (14.0)</td>
<td>36.2 (10.5)</td>
</tr>
</tbody>
</table>

Table 13 shows means and standard deviations of the IMI subscale scores. An ANOVA analyses were used again to assess the differences of users’ engagement and interest among the conditions. The analysis revealed that there was no significant differences in Interest/Enjoyment, $F(2, 138) = 1.88, p = .157$. This was also true in Perceived Choice and Pressure/Tension subscales, $F(2, 138) = 1.84, p = .162; F(2, 138) = 2.11, p = .125$, respectively. However, we have found a marginal significance on Effort subscale, $F(2, 138) = 2.52, p = .083$. This implied that there were marginal significant differences on users’ effort spent while interacting with director agent controlled character. Tukey’s pairwise comparison tests further indicated that the Full and Minimal Guidance conditions are marginally different ($p = .074$). However, Tukey’s test did not reveal any significant differences between the Full and Intermediate Guidance conditions ($p = .820$), as well as the Intermediate and Minimal Guidance conditions ($p = .287$).
### Table 13. Means and SD on IMI raw scores

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Overall IMI</th>
<th>IMI Subscales</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Interest / Enjoyment</td>
<td>Perceived Choice</td>
<td>Pressure / Tension</td>
<td>Effort</td>
</tr>
<tr>
<td>Full</td>
<td>127.3</td>
<td>30.8 (10.7)</td>
<td>28.2 (8.68)</td>
<td>11.9 (5.61)</td>
<td>23.7 (6.41)</td>
</tr>
<tr>
<td>Intermediate</td>
<td>130.7</td>
<td>32.8 (9.17)</td>
<td>29.8 (8.27)</td>
<td>12.6 (5.39)</td>
<td>20.9 (6.43)</td>
</tr>
<tr>
<td>Minimal</td>
<td>122.2</td>
<td>28.5 (10.9)</td>
<td>26.3 (8.56)</td>
<td>14.5 (6.07)</td>
<td>21.8 (7.17)</td>
</tr>
</tbody>
</table>

### 6.3.4.4 Hypothesis IV: Development Time Efficiency

Table 14 shows a qualitative comparison of development time efficiency of the data-driven director agent against a hand-authored approach. The results were measured as rough estimates. For the hand-authoring approach we authored the first slice of the intervention and action DBN models and multiplied the time it took by the total number of slices. For example, since the first slice for the intervention model required almost 2.5 hours to author, and there are 84 slices in the intervention model, the estimated time to author all of the slices is 210 hours.

The results indicate that the hand-authoring approach took twice as much time as the data-driven approach. In addition, the data-driven approach allows new models to be constructed in a much more efficient manner. Typically for machine learning approaches, the time it takes to create revised model is considerably reduced once high-level nodes in the model are set. Small changes within the model can be quickly enacted, in contrast to the hand-authoring approach in which the author has to re-calculate nodes values in each slice to enact the changes.

It is important to note that the hand-authoring effort was performed by the main author of this thesis, who also built the data-driven director agent approach. Therefore, the time recorded in the table is the likely the best-case time for hand-authoring.
### Table 14. Total time on development

<table>
<thead>
<tr>
<th>Tasks</th>
<th>Machine Learning (hrs)</th>
<th>Hand-authoring (hrs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WOZ-enabled CI development</td>
<td>40 hrs</td>
<td>0 hrs</td>
</tr>
<tr>
<td>Corpus Collection</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wizards Training</td>
<td>60 hrs</td>
<td>0 hrs</td>
</tr>
<tr>
<td>Study</td>
<td>49.5 hrs</td>
<td>0 hrs</td>
</tr>
<tr>
<td>Intervention Model Design and Evaluation</td>
<td>150 hrs</td>
<td>20 hrs + 2.5 hrs x 84 slices = 230 hrs</td>
</tr>
<tr>
<td>Model Iteration</td>
<td>10 hrs x 15 iteration = 150 hrs</td>
<td>2.5 hrs x 84 slices x 5 iteration = 1050 hrs</td>
</tr>
<tr>
<td>Action Model Design and Evaluation</td>
<td>180 hrs</td>
<td>30 hrs + 3 hrs x 22 slices= 96 hrs</td>
</tr>
<tr>
<td>Model Iteration</td>
<td>10 hrs x 15 iteration = 150 hrs</td>
<td>3 hrs x 22 slices x 5 iteration= 330</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>779.5 hrs</strong></td>
<td><strong>1706 hrs</strong></td>
</tr>
</tbody>
</table>

### 6.3.5 Individual Differences

In addition to assessing the effectiveness of the director agent decision-making models by evaluating given hypotheses we also explored the role of individual differences in learning and in-game performance during user interactions with the Crystal Island narrative-centered learning environment. There were several areas of individual differences that we have explored, including Identifying differences between high- and low-achieving learners and finding the gender differences in game-play efficiency and learning outcomes.

#### 6.3.5.1 Gender

A series of matched pairs $t$-test were conducted to examine learning differences between male and female in each condition. The results indicate that all Full, Intermediate, and Minimal conditions showed no significance in the learning gains (Table 15). We also conducted a series of ANCOVA analysis by controlling for pre-test to see there are significant differences in learning between the male and female in each condition. The analyses revealed that the male group performed significantly better on post-test than female group in Full Guidance condition, $F(2, 52) = 23.17, p < .001$. This was also true in
Intermediate and Minimal Guidance conditions, $F(2, 45) = 16.25, p < .001$ and $F(2, 44) = 13.64, p < .001$, respectively. This indicates that the male group performed significantly better in learning gains than female group in each condition.

Table 15. Gender differences in learning gains and $t$-test statistics

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Gender</th>
<th>Gain Avg.</th>
<th>STD</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full</td>
<td>male</td>
<td>28</td>
<td>1.21</td>
<td>2.45</td>
<td>1.38</td>
</tr>
<tr>
<td></td>
<td>female</td>
<td>27</td>
<td>1.33</td>
<td>2.91</td>
<td>1.47</td>
</tr>
<tr>
<td>Intermediate</td>
<td>male</td>
<td>24</td>
<td>0.63</td>
<td>2.36</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>female</td>
<td>24</td>
<td>-0.38</td>
<td>2.95</td>
<td>0.40</td>
</tr>
<tr>
<td>Minimal</td>
<td>male</td>
<td>17</td>
<td>1.41</td>
<td>3.55</td>
<td>1.06</td>
</tr>
<tr>
<td></td>
<td>female</td>
<td>30</td>
<td>0.60</td>
<td>2.86</td>
<td>0.68</td>
</tr>
</tbody>
</table>

In addition to the analyses of the learning differences in gender, in-game problem-solving performances were also conducted. Table 16 shows the performance in game-play for each condition. A number of ANOVAs were performed to investigate the differences in time it took between male and female to solve the mystery in each condition. The test revealed that total time it took to solve the mystery was significantly different in Intermediate Guidance condition, $F(1, 46) = 4.30, p < .05$. However, in Full Guidance condition did not show significant differences, $F(1, 53) = 1.50, p = .226$, as well as Minimal Guidance condition, $F(1, 45) = 0.26, p = .612$. A chi-square test performed to indicate the correlation between gender and participants who solved the mystery on each condition. The analyses showed no significant results on each condition, Full Guidance (likelihood ratio, $\chi^2 = 0.32$, Pearson, $\chi^2 = 0.32, p = .57$), Intermediate Guidance (likelihood ratio, $\chi^2 = 1.55$, Pearson, $\chi^2 = 1.51, p = .21$), and Minimal Guidance ($\chi^2 = 0.38$, Pearson, $\chi^2 = 0.39, p = .53$).
Table 16. Gender differences in-game problem-solving performances

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Gender</th>
<th>Solved Mystery</th>
<th>Completion Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>Full</td>
<td>male</td>
<td>28</td>
<td>92.86%</td>
</tr>
<tr>
<td></td>
<td>female</td>
<td>27</td>
<td>96.30%</td>
</tr>
<tr>
<td>Intermediate</td>
<td>male</td>
<td>24</td>
<td>91.67%</td>
</tr>
<tr>
<td></td>
<td>female</td>
<td>24</td>
<td>79.17%</td>
</tr>
<tr>
<td>Minimal</td>
<td>male</td>
<td>17</td>
<td>64.71%</td>
</tr>
<tr>
<td></td>
<td>female</td>
<td>30</td>
<td>73.33%</td>
</tr>
</tbody>
</table>

6.3.5.2 High- and Low-Achieving Learners

The learners were partitioned into high-achievers (above mean) and low-achievers (below mean) based on the mean value of overall learning gain. ANOVA tests were performed on each condition to investigate the differences in time it took between high- and low-achievers to solve the mystery (Table 17). The tests revealed that there was significant difference in Minimal Condition, $F(1, 45) = 8.95, p < .05$. However, in Full and Intermediate Conditions the results showed marginal differences, $F(1, 53) = 3.51, p = .066$ and $F(1, 46) = 3.48, p = .068$, respectively.

To analyze the difference in the number of users who solved the mystery between high- and low-achievers on each condition, a chi-square test was performed. The results showed that the correlation is significant on Full (likelihood ratio, $\chi^2 = 3.40$, Pearson, $\chi^2 = 2.85, p < .50$) and Minimal Guidance (likelihood ratio $\chi^2 = 3.85$, Pearson, $\chi^2 = 3.85, p < .50$), but Intermediate Guidance did not show the significance (likelihood ratio, $\chi^2 = 0.03$, Pearson, $\chi^2 = 0.03, p = .86$).
### Table 17. High- and Low-Achievers in-game problem-solving performances differences

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Learner</th>
<th>Solved Mystery</th>
<th>Completion Time (s)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
<td>STD</td>
<td></td>
</tr>
<tr>
<td>Full</td>
<td>low</td>
<td>29</td>
<td>89.7%</td>
<td>1900</td>
<td>575</td>
</tr>
<tr>
<td></td>
<td>high</td>
<td>26</td>
<td>100.0%</td>
<td>1651</td>
<td>383</td>
</tr>
<tr>
<td>Intermediate</td>
<td>low</td>
<td>26</td>
<td>84.6%</td>
<td>2088</td>
<td>579</td>
</tr>
<tr>
<td></td>
<td>high</td>
<td>22</td>
<td>86.4%</td>
<td>1770</td>
<td>599</td>
</tr>
<tr>
<td>Minimal</td>
<td>low</td>
<td>20</td>
<td>55.0%</td>
<td>2704</td>
<td>318</td>
</tr>
<tr>
<td></td>
<td>high</td>
<td>27</td>
<td>81.5%</td>
<td>2276</td>
<td>576</td>
</tr>
</tbody>
</table>

#### 6.3.5.3 Game Play Experiences

The users on each condition were partitioned using the demographic survey conducted by during the study. The partition is based on the mean value of overall game time users played in a week. A matched pairs $t$-test revealed that only users in the *Full Guidance* condition showed statistically significant learning gains. Users in the *Intermediate* and *Minimal Guidance* conditions did not show significant learning gains (Table 18). A series of ANCOVA analysis is conducted by controlling for pre-test to find the differences between high and low game experience users. Analyses revealed that all conditions indicated significances. For *Full Guidance* condition, $F(2, 52) = 25.76, p < .001$, *Intermediate* condition, $F(2, 45) = 16.25, p < .001$, and *Minimal* conditions, $F(2, 44) = 13.64, p < .001$ revealed strong results.

ANOVA tests were performed on each condition to investigate the differences in time it took between high and low game experience users to solve the mystery (Table 19). The tests revealed that there was significant difference in Intermediate Condition, $F(1, 46) = 14.41, p < .001$. However, in *Full* and *Minimal Conditions* the results did not show the differences, $F(1, 53) = 2.83, p = .09$ and $F(1, 45) = 0.024, p = .88$, respectively. Also, a chi-square test was performed to find the correlations. The results also showed only Intermediate Condition
(likelihood ratio, $\chi^2 = 5.08$, Pearson, $\chi^2 = 4.69$, $p < .05$) showed significance, but Full (likelihood ratio, $\chi^2 = 0.25$, Pearson, $\chi^2 = 0.25$, $p = .62$) and Minimal Guidance ($\chi^2 = 1.13$, Pearson, $\chi^2 = 1.13$, $p = .72$) did not show the statistical significances.

Table 18. Game play experiences differences in learning gains and $t$-test statistics

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Experiences</th>
<th>Gain Avg.</th>
<th>STD</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full</td>
<td>low</td>
<td>0.65</td>
<td>2.96</td>
<td>0.70</td>
<td>.49</td>
</tr>
<tr>
<td></td>
<td>high</td>
<td>1.83</td>
<td>2.28</td>
<td>2.09</td>
<td>&lt; .05</td>
</tr>
<tr>
<td>Intermediate</td>
<td>low</td>
<td>-0.48</td>
<td>2.84</td>
<td>0.62</td>
<td>.54</td>
</tr>
<tr>
<td></td>
<td>high</td>
<td>0.68</td>
<td>2.46</td>
<td>0.77</td>
<td>.44</td>
</tr>
<tr>
<td>Minimal</td>
<td>low</td>
<td>0.64</td>
<td>3.11</td>
<td>0.65</td>
<td>.52</td>
</tr>
<tr>
<td></td>
<td>high</td>
<td>1.18</td>
<td>3.17</td>
<td>1.08</td>
<td>.28</td>
</tr>
</tbody>
</table>

Table 19. Game play experience differences in-game problem-solving performances

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Gender</th>
<th>Solved Mystery</th>
<th>Completion Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>Full</td>
<td>low</td>
<td>96.2%</td>
<td>1901</td>
</tr>
<tr>
<td></td>
<td>high</td>
<td>93.1%</td>
<td>1676</td>
</tr>
<tr>
<td>Intermediate</td>
<td>low</td>
<td>73.9%</td>
<td>2247</td>
</tr>
<tr>
<td></td>
<td>high</td>
<td>96.0%</td>
<td>1663</td>
</tr>
<tr>
<td>Minimal</td>
<td>low</td>
<td>68.0%</td>
<td>2470</td>
</tr>
<tr>
<td></td>
<td>high</td>
<td>72.7%</td>
<td>2446</td>
</tr>
</tbody>
</table>

6.4 Discussion
The findings indicate that users interacting with the CRYSTAL ISLAND story-based learning environment with the embedded machine-learned director agent decision-making models achieved improved learning outcomes. The results showed a significantly improved learning
outcome when the users received the full guidance from the model. However, learning gains for the intermediate guidance version and the minimal guidance version were not found to be significant. Learning differences from the full guidance to the intermediate and minimal guidance were also found to be significant among the conditions. This suggests that users who received full guidance from the tutorial model tended to improve their content knowledge more than users who received minimal or intermediate help.

Interesting findings were also observed concerning users’ in-game problem-solving performance among the conditions. More users who participated in the full guidance condition were able to solve the mystery in an efficient manner than users in the other two conditions. An ANOVA indicates that problem-solving performance was significantly more efficient in the full guidance condition than in either the minimal guidance condition or the intermediate guidance condition. The analyses indicate that users who received full guidance were found to have significantly outperformed the other conditions with respect to learning gains and problem-solving performance.

The findings also suggest that users who were in the intermediate condition did not show much difference in in-game problem-solving performance in comparison with the full guidance condition users, yet they did not improve their learning outcomes. We hypothesize that in the intermediate guidance conditions users were able to receive sufficient help from the model to solve the mystery efficiently; however, they did not receive adequate help from the model to enhance their learning. Additional investigation is needed to better understand the cause.

Unfortunately, findings did not indicate that differences on user engagement and interest with the CRYSTAL ISLAND learning environment among the conditions. Measures for engagement and interest—presence and subject impression questionnaire from intrinsic motivation inventory—were not found to be significantly different among all conditions. Only the effort
subscale in the IMI showed marginal difference between the Full and Minimal Guidance conditions.

The findings of the individual analysis reveal several interesting results. With respect to gender differences, males tend to perform better in learning, as well as in-game problem-solving efficiency. ANCOVA analyses showed that male learners performed significantly better on post-tests than female learners in all condition. Male learners took less time to complete the game in all conditions. It is not evident whether female learners were off-task at this point. However, the findings suggest that male learners were focused on their problem-solving activities than female learners. The results also hold in high- and low-achieving learners. High-achieving learners appears to focus more on their tasks to solve the problem. As an evidence high-achieving learners has significant differences in time it took to solve the mystery when compared with low-achieving learners.
Chapter 7

Conclusion

Interactive narrative environments offer significant potential for crafting engaging story-based experiences that are tailored to individual users. Devising accurate models of director agent decision-making strategies is critically important for creating optimal narrative experiences for users. Although previous investigations have explored techniques for modeling director agents in interactive narrative, little work has resulted in the design of director agent decision-making strategies. This dissertation describes a dynamic Bayesian network-based data-driven director agent decision-making model that has been implemented for a narrative-centered learning environment. The models actively observe users interacting with a storyworld and offer narrative decisions to dynamically guide individual users for an effective narrative experience in a storytelling environment.

7.1 Summary
In this dissertation, we have presented an empirically driven model of director agents’ decision-making strategies for interactive narrative-centered discovery learning environments. A corpus collection was conducted using a Wizard-of-Oz methodology with users interacting with a WOZ-enabled version of an interactive narrative-centered learning environment.
Given the promise that interactive narrative-centered learning environments have shown, we have developed two empirically driven models of direction agent decision-making strategies: an *intervention decision model* (Lee, Mott, & Lester 2011c) and an *action decision model* (Lee, Mott, & Lester 2011a). The intervention decision model determines when the next narrative action should occur. The narrative action decision model determines which narrative-centered narrative actions to perform. Both models were developed using an empirically driven methodology. By utilizing a corpus of human interactions within an interactive narrative-centered learning environment, dynamic Bayesian networks (DBN) were employed to learn the two models of director agent decision-making strategies. The predictive accuracy results indicate that using empirically derived dynamic Bayesian networks for director agent decision-making strategies can make accurate narrative intervention and action decisions.

An empirical evaluation of the machine-learned models of director agent decision-making strategies for real-time interaction with a narrative-centered discovery learning environment was conducted. Three different conditions were employed in the empirical evaluation: *Minimal Guidance, Intermediate Guidance*, and *Full Guidance*. We investigated differences in learning gains, in-game performance, and user engagement and interest during user interactions with the storyworlds. We also investigated the development time efficiency of the data-driven director agent approach against a hand-authoring approach.

It was found that users in a full guidance condition exhibited significant learning gains and problem-solving performance. A detailed analysis of the differences in learning and in-game problem-solving performance among the conditions was also performed. It was found that there were statistically significant differences between users who received full guidance and students who received intermediate or minimal guidance. The findings suggest that machine-learned models of director agent decision-making models can improve learning outcomes and in-game efficiency. A study of authoring efficiency was also undertaken. It was found
that the development time for the hand-authoring approach required twice as much time as the data-driven approach.

In short, the results of the studies are encouraging and suggest that empirically driven models of director agent decision-making strategies using dynamic Bayesian networks appear to be a promising approach to interactive narrative.

### 7.2 Limitation
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. 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 tradeoffs are particularly salient in graphically intensive games, which may have limited CPU and memory resources available for AI-related computations. Also, cognitive models of learning and affect should be employed with the framework. Although the current model can create effective stories, the stories can be more effective by tailoring to the specific individual needs when model corresponds with cognitively guided direction.

### 7.3 Future Work
Several directions for future work are promising. First, it will be important to understand the relationship between learning outcomes and engagement with respect to the behavior of real-time director agent decision-making models for narrative-centered discovery learning environments. Although we did not find significant results among the conditions, an earlier study suggests there is a strong positive correlation between learning outcomes and increased
engagement (Rowe, Shores, & Lester 2010). A follow-on investigation should be conducted to determine whether these results hold for our system.

Another promising direction for future work is exploring the role of individual differences in learning, in-game performance, and engagement during user interactions with the CRYSTAL ISLAND narrative-centered learning environment, in addition to the high- and low-achieving learners and gender. There are other areas of individual differences that should be explored. Understanding more in-depth behaviors among users who had different levels of prior game-play experiences is another possible area of study. The results could lead to improved design via a better understanding of how interactive narrative systems should behave while interacting with users. Customizing narratives to personality traits of users is another potential area of investigation. During our evaluation we utilized the Big Five personality traits questionnaire (Costa & McCrae 1992). A potentially fruitful line of investigation is to identify means for customize narratives for users of particular personality types to achieve improvements in learning, in-game performance, and engagement.

User modeling plays an important role in guided discovery learning environments. By utilizing the user’s knowledge of salient facts in the story, a narrative decision-making model can guide users into pedagogically effective narrative experiences. Within the CRYSTAL ISLAND environment users complete a diagnosis worksheet while solving the science mystery. This tool is utilized to help diagnose the user’s current knowledge and problem solving approach. However, the tool does not represent an accurate view of users’ goals and beliefs. More sophisticated inference techniques, perhaps integrated into narratives themselves, are required to infer user models. More effective techniques for inferring users’ goals and beliefs could contribute to improved user-adaptive narrative experiences.

During the data collection, wizards used natural language dialogue to guide users when unexpected behaviors were encountered. Imbuing characters with sophisticated natural language dialogue capabilities offers a means for guiding users through stories, so devising
adaptive models of narrative-centered interactive dialogue is a promising line of investigation. Another potentially fruitful direction is developing more effective models of affect understanding and affect generation for interactive narrative systems. Equipping director agent decision-making models with the ability to reason about users’ affective states could provide more sophisticated human-centered inferential capabilities that could serve as the foundation for the next generation of interactive narrative technologies.

7.4 Concluding Remarks
This dissertation was motivated by the objective of improving customized narrative and learning experiences for individual users while interacting with narrative environments. We believe that an empirical approach of devising accurate models of director agent decision-making strategies hold much promise for pursuing the goal creating effective individualized narrative experiences. Although the interactive narrative research community has been studying techniques for modeling director agents, it has only recently begun to explore data-driven approaches. The dissertation describes how empirically informed models of director agents can be developed and evaluated with a corpus-based narrative methodology. This work represents a first step toward an empirically driven approach to narrative generation, and we hope it will form the basis for future investigations interactive narrative.
References


Riedl, M., Saretto, C., and Young, M. 2003. Managing Interaction Between Users and


Appendices
Appendix A
Corpus Collection Handout Materials

During the Wizard-of-Oz corpus collection participants were initially given a controlled back-story about Crystal Island, which also includes a character profile sheet listing the character’s names, their photos, and their role in Crystal Island (A1) and a keyboard and mouse control reference sheet (A2). Once the participants completed reviewing the set of Crystal Island handouts, they were given general instructions (A3) that help the data collection process and facilitate efficient interactions with the storyworld.

A.1 Crystal Island Storyworld
The Crystal Island storyworld is situated on a recently discovered volcanic island where a research station (Figure 1) has been established to study the island’s unique flora and fauna. There are several characters in the Crystal Island storyworld: Ford Patterson (zoologist), Kim Lee (camp nurse), Quentin Nash (camp cook), Audrey Newsome (botanist), Alex Reid (player), Bryce Reid (lead scientist), and Al Cunningham (camp foreman). The user plays the role of Alex Reid visiting his/her father, Bryce Reid, who serves as the research station’s lead scientist.
Scenario
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 the camp foreman, Al. Al tells her that Audrey, Ford, and her father were out on an expedition gathering specimens a few days ago. Their expedition was scheduled to last for two days; however, they failed to return to the camp on time. Al found this very unusual since they were always punctual before. Fearful for their safety, Al led a group of people from the camp to search for them. After two-days of searching, they found them fallen ill on the south side of the Island. It appears the group lost their way, became ill, and could not make it back to the camp. They are currently in the infirmary being attended to by the camp’s nurse.

Upon hearing the news, Alex went straight to the infirmary to see her father and his colleagues. Kim, the camp’s nurse, tells her that their condition is not good. Her father seems to be much worse than the others. Kim is baffled by the illness and does not know what could have caused it.

Goal
This Crystal Island scenario features a science mystery in which the user plays the role of a “medical detective.” As members of the research team have fallen ill, it is the user’s task to discover the cause of the disease and its source.

**Location**
In this scenario, all events occur within the infirmary and laboratory, which are housed together in the same building. The user will not explore other areas of the island during this episode. The user will navigate around the environment, conduct tests in the laboratory, ask questions of patients in the infirmary regarding their symptoms, and carry on detailed conversations with the camp’s nurse in order to solve the mystery.

**Interacting Characters**
The user can interact with the nurse, patients, and testing equipment in the lab. The nurse has been treating the patients since they arrived back at the camp and has a good understanding of their symptoms. The user is strongly encouraged to collaboratively work with the nurse to arrive at a diagnosis and cause of the illness. Specially, the user can ask the nurse about patients’ symptoms, their current conditions, or other information that may be helpful. However, the nurse is not a medical specialist for specific diseases and may not know all of the answers. The nurse and user will communicate with each other via spoken dialogue (i.e., voice chat). The user can also interact with patients to get their symptoms.

*The user should follow the scientific method while solving the mystery. It is strongly encouraged that the user communicates with the nurse whenever they have formed a new hypothesis for the type of disease or its cause. The primary goal of this study is for the user to collaborate and communicate with the nurse to correctly diagnosis the illness working together as a team.*
During the user’s exploration of the environment they will have the opportunity to interact with four characters, shown in Figure A.1.2.

Additional Background Information

- According to Quentin (camp cook), the expedition team took two days worth of supplies with them. He thinks they must have supported themselves during the other days by collecting food and water from the island. Also, he found some items in their bags which were not part of the supplies he gave to them for their expedition.
- Kim (camp nurse) asked Quentin to bring the supplies he found in their bags to the lab. Kim also asked him to bring a set of supplies he gave them before they left to the lab.
- Within the laboratory, there is a bookcase that contains information regarding many known diseases including information about their causes and symptoms.

Feel free to use the “Interacting with the Virtual Environment” handout (describing the keyboard and mouse controls) handout, throughout your interaction. If you have any questions at this time, please ask.
A.2 Interacting with Virtual Environment

<table>
<thead>
<tr>
<th>Keyboard Controls</th>
<th>Mouse Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>F</strong> – Worksheet</td>
<td><strong>Left Mouse Button</strong> – Drop object</td>
</tr>
<tr>
<td><strong>W</strong> – Move forward</td>
<td><strong>Mouse up</strong> – Look up</td>
</tr>
<tr>
<td><strong>E</strong> – Use object</td>
<td><strong>Mouse left</strong> – Turn left</td>
</tr>
<tr>
<td>(examine/pickup/drop/operate)</td>
<td><strong>Mouse right</strong> – Turn right</td>
</tr>
<tr>
<td><strong>A</strong> – Move left (strafe)</td>
<td><strong>Mouse down</strong> – Look down</td>
</tr>
<tr>
<td><strong>S</strong> – Move backward</td>
<td><strong>CTRL</strong> – Duck</td>
</tr>
<tr>
<td><strong>D</strong> – Move right (strafe)</td>
<td><strong>Spacebar</strong> – Jump</td>
</tr>
</tbody>
</table>

Figure A.2: Keyboard and Mouse Control Handout

A.3 General Instructions for Interacting with Virtual Environment

List of the following general instructions were given to the participants to improve the corpus collection and interaction in efficiency manner.

- Video is recorded during the session. The video data will be used to devise new learning environment technologies. Any interference with the recording will result incorrect data collection. Please,
  - Do not wear hat
  - Do not chew gum
  - Do not cover your face or mouth with hands during the session
Do not move outside of the computer screen during the session

- During your session you should assume the role of Alex helping to collaboratively solve the science mystery with Kim the camp nurse.
- Kim will refer to you as Alex so please attempt to stay “in character” during your session. Also, communicate with Kim the character not the person who is playing the role of Kim.
- As much as possible, communicate with Kim, the camp nurse, in a “natural” face-to-face manner (i.e., move close to her in the environment and carry on a natural conversation). Although the voice chat system allows you to communicate with Kim regardless of where you are located, do not attempt to communicate with her if she is far away from you (e.g., in another room). Instead move closer to her before communicating.
- Press ‘E’ to examine/pickup/drop/operate objects.
- Press ‘F’ to enter choices in the worksheet. Whenever you gathered vital information for solving the mystery while paying the game, you should update your fact sheet.
- Solve the mystery by following the scientific method.
Appendix B
Wizard Protocol

To improve the consistency of the wizards’ tutorial planning, narrative planning, and natural language dialogue activities, a set of guidelines were given to wizards. They iteratively developed and refined through a series of pilot studies. The resulting protocol included a high-level procedure for the wizard to follow (B1) and narrative decisions that occur while interacting with participants during corpus collection, along with abstract directions and examples what the wizard could do when decision has been chose (B2). The protocol also included information about the symptoms of each of the characters in the storyworld and infection testing results (B3) and short descriptions of the symptoms associated with each of the diseases mentioned in the CRYSTAL ISLAND (B4). The wizards also received all the materials that participants reviewed during the corpus collection.

B.1 Wizard Interaction Guideline

Abstract Procedure
1. Guide the student to look at the scientific method poster.
   • Ask if she knows the scientific method.
   • If yes, ask her to explain how the scientific method works
   • If no, explain briefly about the scientific method process
2. Guide the student over to examine the patients and learn about their symptoms
• Wizard should mention all three characters are suffering from the same disease at the beginning of the story.

• Wizard will receive prompts on screen when the student starts examining a patient and when they close the symptoms dialog.

• Once the student closes the symptom dialog, discuss the patient’s symptoms with the student (e.g., asking what are their common symptoms).

• Ask student to fill out salient symptoms on their fact sheet.

• At this point, the student should know about the patient symptoms and the incubation period (i.e., the team was gone for 4 days).

3. Guide the student over to the library area to read books regarding diseases.
   • Wizard receives prompts on screen when student starts or finishes reading a book.
   • While the student reads the book, pretend to read the same information as student.
   • Once the student finishes reading the book, discuss the contents with the student:
     o Ask about how this disease might be relevant based on the patients’ symptoms and incubation period.
     o Discuss whether the disease could be ruled out or considered as the possible illness.
     o Compare with earlier disease books that have been read. What makes this disease different from the others? Why could this disease be the possible cause of the illness when others have similar symptoms? Or why the current disease could be ruled out when others have similar symptoms but considered as possible causes.

• Wizard should attempt to bring out the pedagogical reasoning from the student on each disease that might be the cause of illness.

• Once all the diseases have been discussed, work with the student to recap what you have discussed.

• Ask student to form a their hypothesis (or what it currently is)
• At the end of this step, the wizard and student should know in detail what diseases have been ruled out and what diseases are still possible, and why.
• Ask student to modify their fact sheet. Ask student to enter what they think is the most probably disease in the fact sheet even thought several are still possible.

4. Take the student over to test the “camp food items” and “outside camp food items”
• Wizard will receive test results for each item as the student tests them
• Discuss with the student after each object is tested. Based on test result,
  o Whether earlier hypothesis should change.
  o If yes, discuss and form new hypothesis.
  o If no, then consider whether the hypothesis should have changed based on the object being the transmission source (e.g., since this object is negative then it’s probably not the source and thus the diseases is less likely) and discuss that with the student. Ask if this changes their hypothesis.
• Wizard should attempt to go beyond just asking questions to the student. Wizard might suggest or point out if the object was mentioned in any of disease books (e.g., perhaps say to the student that you think you remember something about that object being mentioned as a potential transmission source in one of the books).
• Form a new hypothesis if needed
• Ask student to modify the worksheet based on new information and hypothesis.
• Repeat until student solves the mystery (introduce “narrative decisions” as appropriate)

Interaction Guideline
• The main goal of the wizard is to guide and collaborate with the student to solve the mystery
  o Discuss with student all subjects that might help the student solve the cause of the illness.
  o Guide student to use the scientific method approach to find illness cause
• Wizard should act as a companion agent where she is trying to solve the problem with student collaboratively by gathering and discussing evidence together.
• Use narrative dashboard to create narrative events. The dashboard lists a number of narrative decisions that the wizard has to perform during the story.
• Always act and communicate as Kim, the practitioner nurse.
• Kim has limited knowledge about the diseases. She does not have full knowledge to solve the mystery when the session begins. So the wizard needs to act accordingly. After execution of each of the abstract procedure mentioned above, wizard also learns the knowledge that the student does and can use it to assist in solving the mystery.
• Disease related books are in the laboratory.
• When student develops a hypothesis, wizard needs to make sure her approach is constructive and make sense.
• When student draws a conclusion, wizard needs to make sure the conclusion supports her hypothesis and conclusions based on sufficiently analyzed data.
• For reporting result (in the scientific method), ask student to provide an explanation for the steps, which allows her to arrive the current conclusion.
  - How she gathered information and what kind information she gathered.
  - How she formed hypothesis and what was her hypothesis.
  - How she confirmed her hypothesis was valid.
  - How she drew the conclusion.
• Session is finished once nurse agrees with the cause of disease and the worksheet is sufficiently filled out.
• Suggest student should fill out the worksheet. Make sure the student constantly fills out the fact sheet whenever she gathers vital information for solving the mystery.
• Wizard should be with the student, except during performing specific narrative events.
• Wizard should interact and communicate with Alyx, not the student (i.e., the student is playing the role of Alyx).
- As much as possible, communicate with Alyx in face-to-face manner. Do not try to communicate with Alyx if she is far away, even if she is asking a question like you are next to her. Try to move closer to her for any communication.

### B.2 Narrative Decisions Guideline

<table>
<thead>
<tr>
<th>Decisions</th>
<th>Occurrences</th>
<th>Examples (wizard)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduce scientific method</td>
<td>• Event occurs early in the episode.</td>
<td>“Alex, come over here for second. I believe that we should take the scientific method approach to find cause of illness. By correctly define what we need to solve, get information regarding diseases, form hypothesis based on current information, analyze information we have, and maybe we can draw conclusion at that point. So, we can tell what caused them to sick. This poster shows steps of scientific method”</td>
</tr>
<tr>
<td></td>
<td>• Wizard suggests student that the right approach to take for finding the cause of disease is the scientific method.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Wizard briefly explains scientific method by going over to the poster with the student.</td>
<td></td>
</tr>
<tr>
<td>Examine patients’ symptoms</td>
<td>• Event occurs early in the story.</td>
<td>“Alyx, would you like to examine your father, Audrey and Ford’s symptoms?”</td>
</tr>
<tr>
<td></td>
<td>• Wizard suggests student to examine symptoms of Audrey, Bryce, and Ford.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• May occur multiple times in the story.</td>
<td></td>
</tr>
<tr>
<td>Read relevant disease books</td>
<td>• Event occurs after student examined the symptoms of patients.</td>
<td>“Alyx, there are some books in the lab that may help us determine the illness. I believe this will give us more information.”</td>
</tr>
<tr>
<td></td>
<td>• Wizard suggests student that reading books on diseases may provide more information to form a hypothesis.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• May occur multiple times in the story.</td>
<td></td>
</tr>
<tr>
<td>Introduce worksheet</td>
<td>• Event occurs throughout the story.</td>
<td>“Alyx, as you modify your hypothesis and gather new facts, I recommend that you fill out the worksheet. This</td>
</tr>
<tr>
<td></td>
<td>• Whenever student gets new</td>
<td></td>
</tr>
<tr>
<td>Event Type</td>
<td>Description</td>
<td>Student Response</td>
</tr>
<tr>
<td>------------------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Introduce new symptom (headache)</td>
<td>• Event occurs when wizard thinks it is good time to introduce new symptom</td>
<td>“Alyx, it seems they are all suffering from a headache now. They constantly complain about it.”</td>
</tr>
<tr>
<td></td>
<td>• during the interaction with student.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Wizard should visit patients after hearing moaning sound.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Wizard brings new symptom to the student that may help narrow down the disease.</td>
<td></td>
</tr>
<tr>
<td>Perform contaminated object testing (banana)</td>
<td>• Event occurs during the process of testing foods and drinks.</td>
<td>“Bananas tested positive for bacteria. Earlier you said it might be Salmonella. Has this new result changed your hypothesis? What do you think happened?”</td>
</tr>
<tr>
<td></td>
<td>• Wizard discusses with student whether she can draw conclusions from this result or form a new hypothesis.</td>
<td></td>
</tr>
<tr>
<td>Perform objects testing on items the team brought back</td>
<td>• Event occurs when student tests objects on the “second” table.</td>
<td>“Alyx, these are the items that we found in their bag. According to the cook, they are not part of the supplies they took with them on the expedition team. We should test them.”</td>
</tr>
<tr>
<td></td>
<td>• Wizard points out the objects (bananas and egg) to the student that were found in the expedition team’s bag.</td>
<td></td>
</tr>
<tr>
<td>Perform objects testing on items the team took with them</td>
<td>• Event occurs when student tests objects on the “first” table</td>
<td>“Alyx, these are the supply items the expedition team took with them. We should test them.”</td>
</tr>
<tr>
<td></td>
<td>• Wizard points out the objects to the student that they took from the camp for their expedition.</td>
<td></td>
</tr>
<tr>
<td>Introduce definitive new symptom (leg cramps)</td>
<td>• Event occurs when student is having difficult time forming a good hypothesis.</td>
<td>“Alyx, Bryce just complained about having acute leg pains. I think this new symptom will help us reach a diagnosis. What do you think? Can we draw any new conclusions based on this new fact?”</td>
</tr>
<tr>
<td></td>
<td>• Event may not occur if student is heading in the right direction.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• It is up to wizard whether she needs to</td>
<td></td>
</tr>
<tr>
<td>Event Description</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>-----------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Introduce new contaminated object (contaminated water)</td>
<td>Event occurs while student is performing object testing. This object provides critical evidence to form new hypothesis or confirm an existing hypothesis. Wizard brings new item after hearing cook’s voice. Wizard discusses with student about the new evidence. Confirm hypothesis with student or try to help them form a new hypothesis. Wizard and student may draw a conclusion at this point.</td>
<td></td>
</tr>
<tr>
<td>Ensure the worksheet is complete</td>
<td>Event occurs when student hasn’t completed the fact sheet. Wizard reminds student to fill out fact sheet.</td>
<td></td>
</tr>
<tr>
<td>Report story resolution</td>
<td>Event occurs as last part of the story. Once wizard agrees with the student’s conclusion, wizard asks student to explain what might have happened to the patients and how she arrived at such a conclusion via the scientific method.</td>
<td></td>
</tr>
</tbody>
</table>

“Alyx, the camp cook just brought this new item. He said this bottle of water was found one of the team’s bags. Let’s test it.”

“The test result came out positive for infection. What do you think? Does this confirm our hypothesis we developed earlier?”

“Alyx, if you haven’t completed your fact sheet, I recommend we do so now with any new information we recently found.”

“Ok Alyx, could you provide me a scientific method based explanation to make sure we found the right cause of the disease?”

“Great job! I believe we found the cause of disease.”
B.3 Character Symptoms and Object Test Results Guideline

Patients Symptoms
During the student’s interaction in the environment they will learn more about the symptoms from each of the characters. The following information is provided to the student:

<table>
<thead>
<tr>
<th></th>
<th>Audrey Newsome (Botanist)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>“I am doing better than the others. I have diarrhea and vomiting. I am exhausted and dehydrated.”</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Bryce Reid (Lead Scientist)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>“My condition is the worst. I have severe diarrhea, vomiting, and stomach cramps. I also have a fever.”</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Ford Patterson (Zoologist)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>“I am suffering from diarrhea, vomiting, and severe stomach cramps. I am exhausted and dehydrated.”</td>
</tr>
</tbody>
</table>

Infection Test Results
The student sees the following test results as they test objects in the lab. The wizard is provided the results of the test below (and in the game), however, they should rely on the student informing them of the test results via conversation.

Supply items the expedition team took from the camp:
Fresh Bottle of Water – Negative
Ham Sandwich – Negative
Milk – Negative
Orange – Negative
Supply items found in their bags by the cook:
Opened Bottle of Water – Positive
Bananas – Positive
Egg – Negative

B.4 Disease Symptoms Guideline
The following table outlines the symptoms associated with each of the diseases mentioned in this CRYSTAL ISLAND episode. The student learns about these symptoms while reading books in the lab’s library. For the wizard’s part they should use this information to ensure that the student is following the Scientific Method and have properly analyzed the situation. The camp nurse is somewhat familiar with each of these diseases; however, her memory is a bit rusty and consulting the details in the books is necessary to ensure an accurate diagnosis.

<table>
<thead>
<tr>
<th>Symptoms \ Diseases</th>
<th>Botulism</th>
<th>Cholera</th>
<th>E. Coli</th>
<th>Salmonella</th>
<th>Shigellosis</th>
<th>Staph Food Poisoning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diarrhea</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Vomiting</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Tiredness</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abdominal Cramps</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Leg Cramps</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nausea</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Fever</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Double Vision</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blurred Vision</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dropping Eyelids</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slurred Speech</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difficulty Swallowing</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dry Mouth</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Muscle Weakness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gas</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Muscle Aches</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Headache</td>
<td></td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix C

Description: A Dynamic Bayesian Network Model for Narrative Decision-Making

As noted above we propose using a dynamic Bayesian network (DBN) approach to modeling director agent narrative decision-making. DBNs are extensions of Bayesian networks (Russell & Norvig 2009). A Bayesian network (BN), or belief network, is a graphical model, which is represented as a directed acyclic graph. Each node in the model represents a random variable while directed edges connect nodes, which are conditionally dependent. Bayesian networks provide a compact representation of the full joint probability distribution among the random variables and support probabilistic inference. Figure C.1 shows an example Bayesian network that might be used to model how a character’s health impacts their behavior and appear. Within this network the Exercise and Fever nodes are conditionally dependent on the Sick node and they both directly influence the Sweat node. Intuitively the network is modeling the fact that if someone is sick then their exercise activity decreases while the chance of having a fever increases along with how exercise and fever affect sweating. Additionally the figure shows that every node has an associated conditional probability table.
(CPT) defining the probability distribution. The CPT parameters are typically constructed by a domain expert or learned from data.

![Example Bayesian network](image)

**Figure C.1: Example Bayesian network**

Although Bayesian networks can model events in a natural and efficient manner, they assume that the state of the world is static. They are typically used for problem domains in which the state of the world does not have temporal dependencies since they do not provide a mechanism for representing the time dimension. Interactive narrative is a time-based phenomenon. Director agents utilize numerous observations that change over time to be able to as accurately as possible select the most appropriate narrative decision. Dynamic Bayesian networks (DBNs) can explicitly characterize models’ belief state over time. DBNs provide a natural representation for describing worlds that change dynamically over time (Dean & Kanazawa 1988; Russell & Norvig 2009). A DBN is a directed acyclic graph that incorporates *time slices*, where each time slice contains its own state variables. Figure C.2
shows an example DBN depicting three time slices. It illustrates that at time $t$ the probability of $Sick_t$ is directly influenced by $Sick_{t-1}$ and that $Sick_t$ directly influences the probability of $Sick_{t+1}$. The probability of $Fever_t$ is also influenced by $Sick_t$ node. By utilizing time slices, DBNs support probabilistic inferencing about events that change over time.

**Figure C.2: Example dynamic Bayesian network**

### C.1 Model Structure

In this work we are modeling director agents’ narrative decisions, which change throughout the course of an interactive experience. To support this, we introduce an abstract DBN model for narrative decisions with a network structure as shown in Figure C.3. Three time-slices are illustrated in the figure with the narrative decision from the previous time slice, $D_{t-1}$, influencing the current narrative decision, $D_t$. Within each time slice, observations from the story world, $O_t$, also influence the narrative decision. These observations might include items such as the location of the user or other characters in the virtual world or features of the unfolding narrative.
More formally, we can describe the model of narrative decisions such that it consists of the sequence of narrative decision variables \( D_1, \ldots, D_t \) and sequence of possible narrative observation variables \( O_1, \ldots, O_t \), where \( t \) is the current time slice. With these components we can form the full joint probability distribution over narrative decisions and narrative observations for this network, as follows:

\[
P(D_1, \ldots, D_t, O_1, \ldots, O_t) = \prod_{i=1}^{t} P(D_i | D_{i-1}) \prod_{i=1}^{t} P(D_i | O_i)
\]

where, \( P(D_i | D_{i-1}) \) specifies time dependencies between the narrative decisions and \( P(D_i | O_i) \) specifies dependencies on narrative observations at time slice \( i \).

Figure C.3: Abstract dynamic Bayesian network model for narrative decision-making

Given the full joint probability distribution of the proposed network, we can compute the probabilities of any query variables given some observed evidences in the network. For example, we could compute the probabilities associated with \( D_t \) after setting evidence to determine the most probable narrative decision at time \( t \). However, it is computationally very
expensive. Fortunately, there are several methods that support efficient ways of computing the probability distribution of query variables.

**C.2 Inference**

Bayesian inference, also referred to as belief update, is a process of computing joint distribution probabilities of query variables given some observed evidence variables. For example, the network mentioned above can be used to find the probability distribution of a narrative decision at time $t$ when a set of narrative observations and prior narrative decisions is observed as evidence variables. General algorithms for inference in Bayesian networks have been shown to be computationally complex, in worse case they are NP-hard (Cooper 1990). A number of efficient algorithms have been studied. Computing the exact answer to a given query variable and techniques that approximate the exact answer are two common approaches. For exact inference, techniques such as the *junction tree algorithm* (Lauritzen & Spiegelhalter 1988; Cecil & Adnan 1996), *frontier algorithm* (Zweig 1996), and *interface algorithm* (Murphy 2002) are commonly used. Although exact inference techniques are widely used, as the network size increases they become computationally inefficient. Approximate inference techniques are utilized in large networks. Techniques such as the *Boyen-Koller algorithm* (Boyen & Koller 1998) and *factored frontier algorithm* (Murphy & Weiss 2001) are used when faster results are necessary in larger networks. For our inference task we use the GeNIe/SMILE (Druzdzel 1999) implementation of the *singly connected junction tree algorithm* (Cecil & Adnan 1996) for exact inference in the dynamic Bayesian networks.

**C.2 Learning**

One of the advantages of Bayesian networks is utilizing a corpus of data in their construction. The process of learning from data takes different forms in terms of whether the network structure is known and whether the data is observable (complete data) or hidden (incomplete data). There are two methods for learning in Bayesian networks: *parameter learning* and
structure learning. Structure learning methods learn the directed acyclic graph structure from a data set. With structure learning, the network itself is learned from data. With parameter learning, the network structure is specified by a domain expert and learning only involves inducing the conditional probabilities. In our story-based dynamic Bayesian networks, the network structure was hand constructed. Using a collected corpus, the values in the conditional probability tables for each node in the network were learned.

A general approach for parameter learning is Expectation-Maximization (EM) algorithm (Bilmes 1998). EM is especially useful when data is partially observed because of the missing, hidden, or noisy data. It is an approximation method and finds locally optimal solutions. EM handles the problem by iterating between the expectation (E) step and the maximization (M) step. Each E-step estimates expectations over the missing or noisy data. In the M-step the maximum likelihood parameter is computed using the expected parameters obtained in the E-step. For our models the EM algorithm from the SMILearn (Druzdzel 1999) library was utilized to learn the CPT parameters.
Appendix D
Student pre/post-test

Please answer each of the following questions to the best of your ability.

1. Which of the following statements is TRUE:
   a. All diseases caused by bacteria have the same set of symptoms
   b. All diseases caused by bacteria have the same incubation period
   c. Some diseases caused by bacterial infections can be transmitted by contaminated drinking water
   d. None of above

2. Your friend began to feel sick this morning and is showing the following symptoms: stomach cramps, fever, and severe diarrhea. She suspects that the source of her illness was some suspicious-looking hamburger meat she ate yesterday. Which of the following diseases is she likely suffering from?
   a. Botulism
   b. Samonella
   c. E. Coli
   d. Cholera

3. The scientific method usually begins with ...
   a. a conclusion
   b. a question
   c. an experimentation
   d. a hypothesis
4. A hypothesis is ...
   a. a process of experimentation
   b. an educated guess based on observations
   c. a random thought
   d. an experiment

5. Which of the following is NOT a symptom?
   a. Fever
   b. Exhaustion
   c. Headache
   d. Shigellosis

6. Which of the following statements is TRUE?
   a. All diseases caused by bacteria are water-borne
   b. All diseases caused by bacteria are food-borne
   c. All diseases caused by bacteria can be either water-borne or food-borne, but not both
   d. None of above

7. What do you do to test a hypothesis?
   a. Guess
   b. Create a spreadsheet with data
   c. Publish a scientific paper
   d. Design an experiment

8. Which of these words would you NOT associate with the scientific method?
   a. Data
   b. Disorganized
   c. Analysis
   d. Orderly

9. Which of these would be called results from the scientific method?
   a. Wondering why people are sick
   b. Discussing symptoms of patients
   c. Identifying transmission source of a disease
10. After gathering information in the scientific method the next step is to ...

   a. form a hypothesis
   b. define a question
   c. draw conclusion
   d. report results

11. What is the incubation period associated with an infectious disease?

   a. Amount of time after symptoms appear before an infected person recovers
   b. Duration between infection with a virus or bacteria until symptoms appear
   c. Duration of time when a virus or bacteria multiplies within an infected person
   d. Amount of time before a diagnosis can be made for an infected person

12. Which of the following is a water-borne infectious disease?

   a. Botulism
   b. Salmonella
   c. E. Coli
   d. Cholera

13. Which of the following does NOT assist in identifying an infectious disease?

   a. Treatment
   b. Incubation Period
   c. Transmission
   d. Symptoms

14. You are part of a team of scientists investigating an illness that may have been caused by contaminated food. Your team has asked the question, "What food source could be causing the illness?" What step should be taken next?

   a. Formulate a hypothesis about the source of the illness
   b. Gather information about the spreading illness
   c. Perform a test to identify the food source
   d. Report findings about the cause of the illness
15. Your friend began to feel sick just a few hours after she ate a sandwich for lunch and is showing the following symptoms: stomach cramps, headache, and vomiting. Which of the following diseases is she likely suffering from?

a. Botulism  
b. Cholera  
c. Salmonella  
d. Staph

16 – 20. Imagine that you are a scientist trying to determine the source of illness that may have been caused by contaminated food. In what order would you complete the following steps?

a. I would examine patients to learn theirs symptoms  
b. I would think about whether I understood the task and define the task question  
c. I would test the food that patients ate  
d. I would make an educational guess on the source of illness based on what I've found  
e. I would make an observation and decide whether my approach would work or whether I needed to change something