Suspenser: A Story Generation System for Suspense

Yun-Gyung Cheong, *Member, IEEE*, and R. Michael Young, *Senior Member, IEEE*

***Abstract*—Interactive storytelling has been receiving a growing attention from AI and game communities and a number of computational approaches have shown promises in generating stories for games. However, there has been little research on stories evoking specific cognitive and affective responses. The goal of the work we describe here is to develop a system that produces a narrative designed specifically to arouse suspense from its reader. Our approach attempts to create stories that manipulate the reader's suspense level by elaborating on the story structure that can influence the reader's narrative comprehension at a specific point in her reading. Adapting theories developed by cognitive psychologists, our approach uses a plan-based model of narrative comprehension to determine the final content of the story in order to manipulate the reader's suspense. In this paper, we describe our system implementation and empirical evaluations to test the efficacy of this system.**

***Index Terms*—Cognitive model, discourse planning, narrative generation, suspense.**

I. Introduction

NARRATIVE has long been associated with games. While most games employ linear narrative as a back story only to engage the gamer with gameplay, classical adventure games and some recent action games such as *Heavy Rain* [1], *L.A. Noire* [2], *Bioshock Infinite* [3], and *Fallout: New Vegas* [4] are designed to offer the player with high narrative experience as well. A number of story generation systems and AI techniques have been developed for creating highly interactive stories [5]–[9]. And yet, there has been little attention to affective properties of narrative, which are, in fact, fundamental to its appreciation by the user. The feeling of suspense, surprise, and curiosity is generally expected in reading and viewing narrative forms [10]–[12]. Among these emotions, suspense has drawn much attention from psychologists due to its significant impact on the reader's enjoyment. In the empirical study conducted by Brewer and Lichtenstein [10], the participants reported that suspense is cardinal for discerning a story from a mere series of events and expressed high satisfaction about their narrative experiences when suspenseful events were presented in the stories.

Our approach addresses the problem of creating suspense in narrative, which keeps the user engaged in the various plots, giving them high entertainment value. We view suspense as the feeling of excitement or anxiety that audience members feel when they are anticipating the occurrence of some event and are uncertain about the event's outcome [13]–[15]. Our work focuses on the class of suspense associated with the perceived likelihood of undesirable outcomes over preferred outcomes. More specifically, implicit in much of the work we present here is the notion articulated by Gerrig and Bernardo [16] in which they view a story's audience as active problem-solvers. In their model of narrative comprehension, a reader's feeling of suspense is affected by the number of potential solutions she can anticipate for the dilemma faced by the protagonist.

To generate suspenseful stories, we set out a basic approach built on a tripartite model, adapted from narrative theory, that involves the following elements: the *fabula*, the *sjuzhet*, and the *discourse* [17]–[20]. A *fabula* is a story world that includes all the events, characters, and situations in a story. A *sjuzhet* is a series of events selected from the fabula and an ordering over those events indicating the order in which they are to be presented to readers. In our approach, the fabula and sjuzhet are represented as plan structures. The final layer, the *discourse*, can be thought of as the medium of presentation itself (e.g., text, film). In this paper, we present Suspenser, a framework that determines narrative contents (i.e., sjuzhet) from a given story world (i.e., fabula) intended to evoke suspense on the part of the reader. The rest of this paper is organized as follows. Section II reviews relevant work in narrative psychology and AI. Section III describes our tripartite model for story generation. Section IV details the Suspenser framework, followed by our evaluation that assesses the performance of Suspenser in Section V. The last chapter concludes with a discussion of the limitations of our system and the plan for future work.

II. Related Work

Theories in psychology and cognitive science view the suspense in the reader's mind as related to diverse narrative elements. The reader's *disposition toward protagonists or outcomes* is a primary condition for her to feel suspense [13], [21]; the reader should be able to anticipate the potential for *conflicting outcomes of an event* [12]. The reader's suspense would increase when she perceives *likelihood of undesirable outcomes relative to preferred outcomes* [14], [15], [22]–[26]. *Prolonging the length of time* the reader anticipates an event or outcome that is potentially harmful to a protagonist can also heighten her suspense [26]. *Discrepancies between the knowledge of the story world held by characters and that held by the reader* can be used as a device to induce the reader's suspense [11], [12], [23]. In the well-known emotion model OCC [13], suspense could be regarded as a specific type of *prospect-based emotion*, one which is evoked when an individual anticipates the occurrence of events whose outcome is uncertain. Their account of suspense

Manuscript received July 10, 2013; revised February 03, 2014; accepted May 06, 2014. Date of publication May 14, 2014; date of current version March 13, 2015. This work was supported in part by Award 0092586 and Award 0414722, and the EU FP7 ICT Project SIREN (Project: 258453).

Y.-G. Cheong is with the College of ICE, Sungkyunkwan University, South Korea (e-mail: ygcheong@gmail.com).

R. M. Young is with the Department of Computer Science, North Carolina State University, Raleigh, NC 27695 USA (e-mail: young@csc.ncsu.edu).

Digital Object Identifier 10.1109/TCIAIG.2014.2323894

involves hope and fear; a person hopes his favorable consequence will be realized while he fears the occurrence of undesirable consequences. Gerrig and Bernardo [16] hypothesize that a reader's feeling of suspense is affected by the number of potential solutions available for the protagonist. Under this model, an audience will feel an increased measure of suspense as the number of options for the protagonist's successful outcome(s) decreases. To confirm this hypothesis, Gerrig and Bernardo performed a number of experiments with human subjects. The subjects were provided different text versions of a story where a protagonist is in danger and tries to escape. The various versions of the story differed in the number of solutions available to the protagonist. After reading the text, subjects were asked to rate their estimation of the likelihood of the protagonist's escape as well as their suspense levels. The data from the experiments showed that the readers reported higher suspense as possible solutions were eliminated.

Quite a number of interactive storytelling systems have been implemented so far [5]–[9], [27]–[31]. In particular, MINSTREL [31], a case-based approach to modeling human creativity, is most comparable to our approach. At the core of the MINSTREL design is a transform-recall-adapt process. On receiving a problem specification as input, MINSTREL retrieves a case from its memory that is similar to the problem given as input. If the case is identical to the problem, the original solution is used. If the case differs from the input problem, the original solution is modified. As an end-to-end system, MINSTREL extensively attempts to solve multiple story-generation issues such as themes, coherency, characterization, tragedy, suspense, and foreshadowing. In MINSTREL suspense is created by fabula-level generation, relying on the psychological evidence that readers feel more suspense when they strongly care about the character [13], [21] and when the presentation of a significant outcome is prolonged [26]. In our approach suspense is created by sjuzhet-level generation, focusing on selecting relevant contents when a fabula is given as an input.

The advantage of using Suspenser is that it can work with an already existing fabula generator such as IPOCL [32] which generates a story that is coherent with each character's intent or CPOCL [33] that plans a story containing conflict. The drawback of Suspenser is that it cannot add a new story event that is not present in the fabula. For instance, imagine a scene that an agent is climbing on a high-rise building with special gloves on. To make the situation more suspenseful, one of his glove gadget runs out of battery and let him fall. Suspenser cannot add such an event that endangers the protagonist at the sjuzhet level without consulting with the fabula generator. On the other hand, Minstrel can create such an event to heighten suspense. However, Minstrel needs to have a case that instructs how to generate necessary events in a particular situation.

III. A Tripartite Architecture of Story Generation

Following the recent research in narratology claiming the benefits of three-layer model in story analysis [17]–[20], our architecture is designed as a three-stage pipelined architecture which consists of fabula generator, sjuzhet generator, and discourse generator. The fabula is represented as a plan data structure, as in *Definition 1*, created in response to a story request specifying the initial and goal states of the story world. To generate the fabula plan, we use an implementation of the hierarchical, partial-order causal link planner Longbow [34], [35][1], which is discussed in Section IV-A. A fabula (illustrated in Fig. 1) is sent as input to the second component in the pipeline, Suspenser in our work.

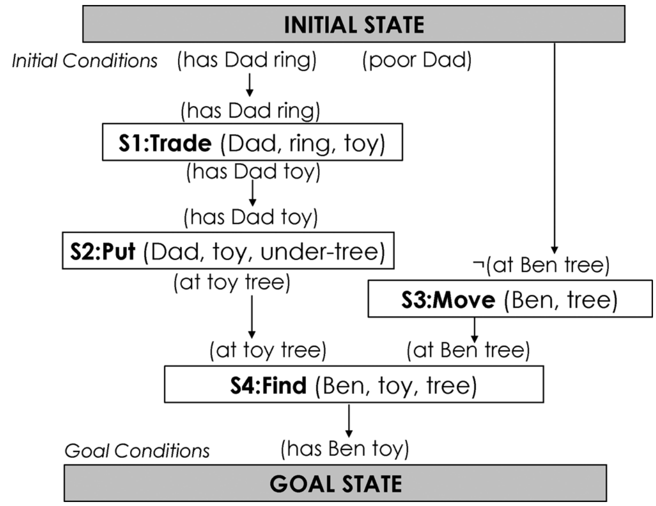

Fig. 1. Example fabula plan structure, illustrating a father getting a toy for his seven-year old son Ben as a Christmas gift. In the diagram, temporal ordering constraints (O) proceeds from the top to the bottom. Rectangles represent a series of plan steps (S), with each action's preconditions enumerated above its rectangle. An arrow between two actions indicates a causal relationship that holds between the two (C), meaning that the effects of the action at the starting point of the arrow establishes a precondition for the action at the arrow's end point. Given an initial state (i.e., the father is poor and he has a ring), this plan is constructed automatically to achieve the goal *(has Ben toy)*. The condition $\neg(at\ Ben\ tree)$ of *S3:Move* action is true under the closed-world assumption: a condition not specified in the initial state is assumed to be false. The plan can be described in text as: "A poor father traded his wedding ring for the toy that his son Ben wants to have. He then put the toy under their Christmas tree. The next day Ben walked to the tree and found the toy that his father left."

Definition 1 (Fabula): A fabula F is a tuple $< S, B, O, C >$ where S is a series of plan steps, B is a set of binding constraints on the variables in steps in S, O is a set of temporal ordering constraints over steps in S, C is a list of causal links between steps in S. Each element of S is a step representing an instantiation of a plan operator contained in a plan library. A plan operator *op* is a tuple $< N, P, E >$ where N is a unique name for the operator, P is a set of preconditions, terms representing just those conditions that must hold for operator to be able to occur, and E is a set of effects, terms denoting just those conditions that change as a result of the action's successful execution. A causal link between two steps s_i and s_j, indicated by the notation $s_i \rightarrow_p s_j$ indicates that s_i establishes as an effect some condition p that is used to establish the precondition of step s_j.

From this fabula, Suspenser constructs a sjuzhet as defined in *Definition 2*. Upon receiving the story structure from Suspenser, the discourse generator produces surface structure (i.e., text, animation). The discourse generator in our current system uses a

[1]In this work, hierarchical planning functionality was not used.

template-based approach which maps a plan step into a single sentence. Suspenser serves as a sjuzhet generator in our work.

Definition 2 (Sjuzhet): A sjuzhet Z is a tuple $< F, S, O >$ where F is a fabula, S is a subset of the plan steps of F to be presented to the user, and O is a set of temporal ordering constraints over steps in S. When O is empty, Z uses the ordering and binding information of F.

We sketch selected narrative-planning models that can serve as a component for each layer. The first layer, the fabula generator, can be replaced by a number of systems that enrich the story world by populating story characters and events. Riedl and Young [32] have developed Fabulist, a story generation system using an Intent-driven Partial-Order Causal Link planner (IPOCL). Fabulist ensures that the actions in the generated story are consistent with characters' intention. In this manner, Fabulist can maintain the balance between plot coherency and character believability. Further, an extension of the IPOCL algorithm is made by [33] to develop a story planner which can create conflicts in narrative. Unlike conventional planners where threatened causal links are strictly prohibited, Conflict Partial-Order Causal Link (CPOCL) allows them under the condition that those events threatening causal links are not executed. Cavazza *et al.* [5] have developed a story generation system that builds a storyline by modeling interactions between autonomous agents using Hierarchical Task Network (HTN) planning techniques. HTN planning represents a plan as a collection of possible subtasks to achieve a higher level goal. The system has been further extended to generate stories that can express the emotions felt by characters through the use of emotional operators which contain mental states as preconditions and effects [36], [37].

As a sjuzhet generator there are some systems that construct an output story by manipulating story events and temporal orderings. [38] describes a reasoning-based approach that assembles two distinct versions of one story that share climax into a single twisted story. Bae and Young [39] have investigated the use for telling events out of chronological sequence focusing on telling future events earlier than their occurrences (*foreshadowing*) or later than their occurrences (*flashback*) aiming at reader's surprise arousal.

Discourse generation relates to different media. For narrative prose generation, Callaway and Lester [40] have developed the AUTHOR system that performs narrative segmentation and chooses appropriate pronoun for a concept, and makes lexical choices that give variation to repetitive expressions. Their experimental study supports that making decisions taking into account the discourse history has a significant impact on the narrative prose quality. In 3D-virtual environments, Darshak, the camera planning system [41] takes as input the story events along with communicative goals and outputs a discourse plan integrating the camera directive actions into the input story plan. The selection of camera actions are made to ensure that the given communicative goals are achieved.

IV. A Computational Model of Sjuzhet Generation for Suspense

Suspenser takes four elements as input: a fabula, a suspense level (either high or low), a goal, and a given point t in the story plan. Then Suspenser determines the sjuzhet, the content to be used to convey the story up to t to a reader for the suspense level. Suspenser consists of three components: the skeleton builder, the suspense creator, and the reader model (see Fig. 2). The skeleton builder identifies a series of *kernels* [42], [43] in the story —important events in a story that cannot be eliminated without harming the story's understandability. Those events that are not included in the skeleton are defined as *Satellites*, events that enrich or elaborate upon the kernels and can be omitted without damaging the storyline. The reader model takes the sequence of kernels and checks them for coherency, essentially modeling a human reader's process of story comprehension to determine what mental model of the story a reader would form, as described later. The sequence is then passed to the suspense creator that again uses the reader model to predict which story elements from the sjuzhet can serve to contribute to manipulate suspense.

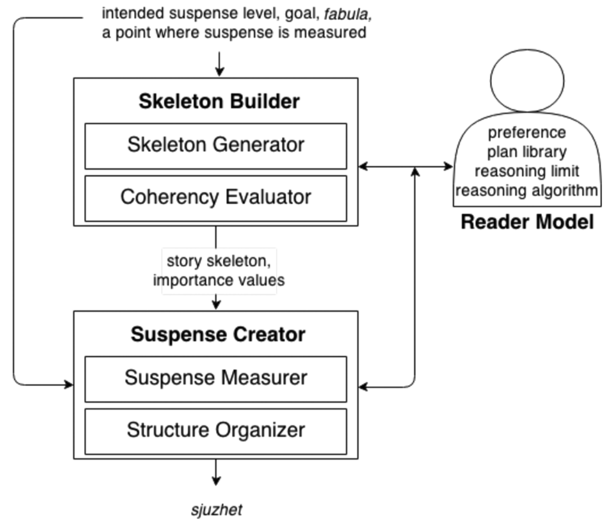

Fig. 2. Suspenser architecture.

A. The Reader Model

The reader model—acting as a proxy for an individual reader's comprehension processes—contains four elements: a reasoning algorithm, a limit on reasoning capacity, a plan library indicating the reader's background knowledge, and a set of preferences that model a reader's preferred types of story elements. For simulating the human *reader's reasoning process* in this paper, we use Crossbow, a C# implementation of the hierarchical partial-order causal link planner Longbow [34], [35]. Prior work has provided strong evidence that types of human task reasoning are closely related to partial-order planning algorithms [44] and that refinement search [45], the type of plan construction process performed by Crossbow, can be used as an effective model of human plan reasoning processes [46]. As a form of *reasoning limit*, an integer counter is used to constrain the number of nodes to be searched during the planning process. To represent the *reader's knowledge*, we use a plan library, consisting of a set of plan operators; each operator consists of a unique name, a set of preconditions and effects, and a set of variables to be instantiated in the planning process. The preconditions of an action in a plan represent just those conditions that must hold for the action to be able to

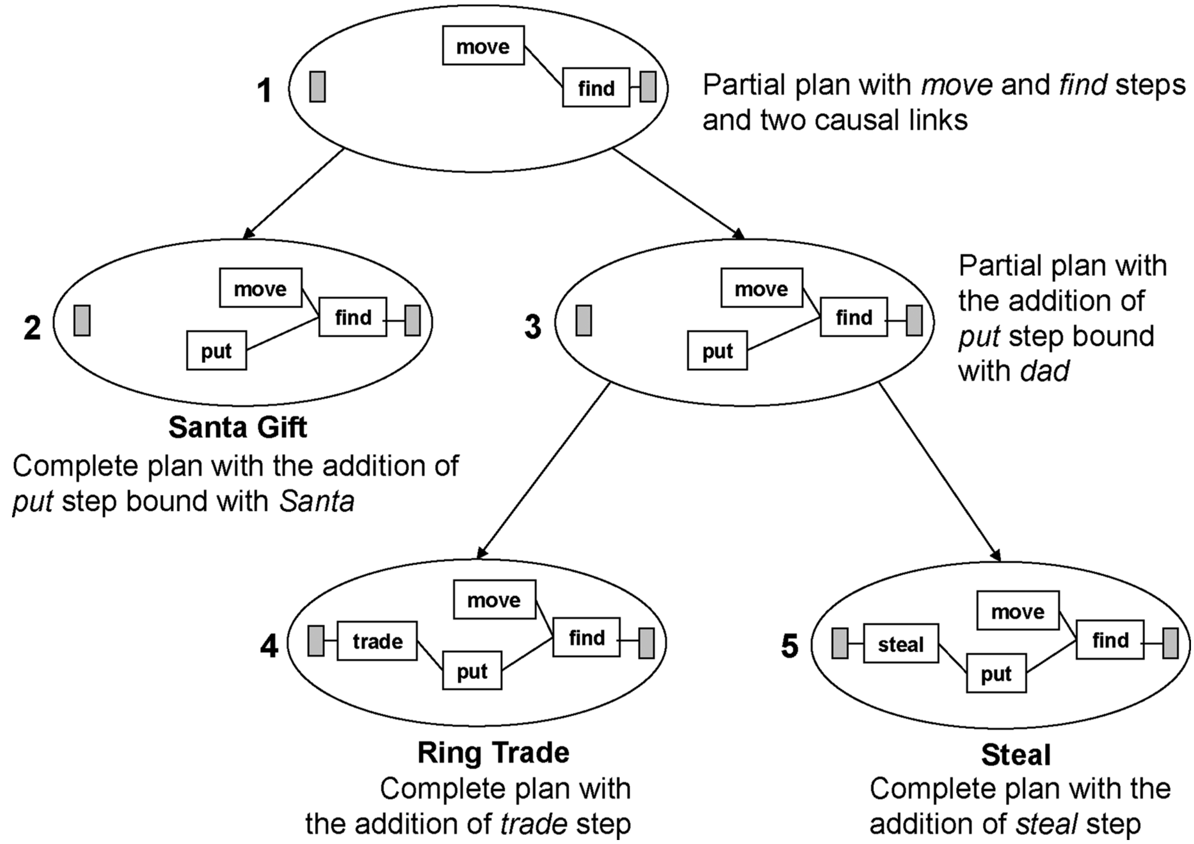

Fig. 3. Plan space graph modeling the reader's forward inference process. The root node #1 represents a partial description of a story's plot as a partial plan which consists of *move* and *find* actions along with causal links. Each arc between two nodes indicates an inference by the reader that reconstructs a single aspect of the story's missing detail as a plan structure. This inference process here leads to the construction of three different complete plans (#2, #4, #5). The node #2 is a complete plan that refines the parent node by adding the *put* action bound with *Santa* as the agent of the action, while the node #3 denotes a partial plan refining the parent node by adding the same *put* action bound with *Dad* as the agent. To satisfy the open precondition of *(has Dad toy)*, two complete plans in the node #4 and the node #5 are generated by adding different actions *trade* and *steal* respectively. The plan in the node #4 corresponds to the fabula plan in Fig. 1.

happen while the set of effects denotes just those conditions that change by the action's successful execution. The *preference function* models the reader's heuristic function used during the refinement search performed as a part of the planning process. If the reasoned story has no break in causal relationship between one event to the next, then it will be understood as coherent.

When a partial description of a story is given, the reader may infer a complete story by filling the gaps to understand it as a coherent story to achieve the protagonist's mission. In refinement search [45], the planning process is modeled as a search through a space of plans which is represented as a directed acyclic graph of partial plan nodes. An arc from one node to another indicates that the plan at the child node is a version of the plan from the parent node with a single plan flaw repaired. In our approach, the root node of the graph is a partial plan taken from the skeleton builder or the suspense creator. Interior nodes in the plan graph represent partial plans with flaws, while leaf nodes of the graph are either complete plans without flaws or plans with flaws that cannot be repairable due to inconsistency in the plan. In plans created by Crossbow, a flaw is either a precondition of some step that has not been established by prior step in the plan, or a causal link that is threatened (i.e., undone) by the effect of some other step in the plan. When adding a new node to the plan space graph, Crossbow creates the child node based on the repair of a single flaw from the parent node. When the flaw is an open precondition, a causal link is established in the new node's plan from either an existing step in the plan or an instantiated operator in the plan library which has an effect that can be unified with the precondition; in the second case, the instantiated step is also added to the parent plan. When the flaw is a threatened causal link, a temporal constraint (i.e., either demotion or promotion) to resolve the threat is added or binding constraints are added to separate the threat involved steps so that no conflicts arise. Fig. 3 shows a plan space resulting from expanding partial plan #1 into three different complete plans (#2, #4, #5) by refining open precondition flaws in parent nodes.

B. *The Skeleton Builder*

Suspenser's primary task, selecting which story elements to tell, eliminates some of the story plan's elements from the discourse describing the story. As more and more elements are excluded from the discourse, however, the resulting gaps in the plan may make the underlying fabula difficult to identify. For example, a story that leaves out the event of Cinderella losing her shoe would not be readily recognized as the well-known version of the Cinderella story. To maintain the identity of the input story, our approach selects a set core events of the fabula, *kernels*, for inclusion in the sjuhzhet using techniques that exploit results from narrative comprehension studies.

The skeleton generator rates the importance of each event based on a method devised by Trabasso and Sperry [47] for extracting important actions that are likely to be included in the story recall. To determine an individual story event's quantitative importance, their approach counts the number of causal relationships with other events. Further, they measure each event's qualitative importance by analyzing its role in the causal chains. Causal chains are a series of events in the story that are causally related. Causal chains contain actions that can be characterized as either opening events, closing events, or continuing events. Opening events introduce characters and the setting and initiate the story. Closing events determine whether the protagonists' main goals are achieved or not. Continuing events causally connect opening events to the closing events in the story.

Drawing on their approach, calculating the importance of each event in a plan was formalised as shown in Heuristic Function 1. Our system approximates an event's quantitative importance by counting the number of its incoming and outgoing causal links. For measuring an event's qualitative importance, we define importance categories: opening-act, closing-act, motivated-act, and contiuing-act types. Opening-acts are the first actions in the story —those that connect propositions from the initial state to later events in the text; Closing-acts are the last actions that occur in the story; motivated-acts are the plan steps that are in causal relationship with preconditions of the goal state; Continuing-acts are a default type since every step in a complete plan is causally related.

1) Heuristic Function 1 (Importance of an Action): $w(a,p)$ returns the importance of an action a in plan p

$$h(a,p)=\frac{k_i In(a,p)+k_j Init(a,p)+k_o Out(a,p)+k_c cc(a,p)}{DistEffect(a,p)}$$

where $In(a,p)$ returns the number of incoming causal links to a coming from steps of the plan p other than the initial step, $Init(a,p)$ returns the number of incoming causal links to a from the plan's initial step, $Out(a,p)$ returns the number of a's outgoing causal links, $cc(a,p)$ returns the causal-chain value of a in the plan, and $DistEffect(a,p)$ returns a value associated with the causal distance between the step a and the goal step of the plan p.

All scaling factors in the function (k_i, k_j, k_o, k_c) are constrained to be real numbers no less than 0. In the formula, the causal-chain value of an event (that $cc(a,p)$ returns) is determined by the event's act type. Important categories (i.e., motivated-acts) are assigned high integer values to give increased likelihood for those acts to be included in the skeleton. Less important categories (i.e., continuing-act) are assigned low integer values. The assigned values are determined empirically through informal experiments. When a step has multiple act types, the type with the highest act value will be chosen. The definition of the normalization function $DistEffect(a,p)$ is informed by a psychological distance effect, which indicates that an action in an episode is more readily comprehended when it is nearer to the episode goal [48]. Foss and Bower [48] define the distance from an action to a goal as the number of actions interposed between them in a subgoal hierarchy constructed in the reader's mind. In our system, the distance from an action to the goal is defined as the minimum number of causal links that relate an action to the goal in a plan. This condition requires the causal distance magnitude ordering of action-goal pairs to be retained in $DistEffect(a,p)$ pair magnitudes, as well. For example, in Fig. 1 the causal distance between the action *S4:Find* and goal step is 1, and the distance between the action *S2:Put* and the goal is 2, which makes the distance of the *action S4-goal* pair nearer than that of the *action S2-goal* pair. Therefore, a function suitable for $DistEffect(a,p)$ would yield a value for the *action S4-goal* pair smaller than that for the *S2-goal* pair.

Once each event's importance is computed based on Heuristic Function 1, the skeleton generator selects the k most important events in the story as *kernels*. The integer value of k, a story's desired length, in our experiments was chosen by a human story writer. Once a skeleton is built, the coherency evaluator tests whether the skeleton is coherent from the reader's perspective as an integral story using an algorithm which is a cycle composed of two phases. The first step uses the reasoning algorithm in the reader model to find complete plans to achieve the protagonist's goals which are consistent with the skeleton candidate. If such a plan is found, the story skeleton is coherent and the program exits. Otherwise, a satellite event in the fabula with the highest importance value is chosen and added to the candidate. Then, the first phase begins again.

C. The Suspense Creator

The suspense creator takes as input the story skeleton and importance values of satellites (see Fig. 2). The central task of the suspense creator is to find a series of additional plan steps α from a portion of the story preceding the step t, which enables the reader to infer the minimum number of solutions. The suspense creator consists of two subcomponents: the structure organizer and the suspense measurer. The structure organizer uses the skeleton as the initial contents of the sjuzhet and then it augments the skeleton with additional story elements that would result in heightening suspense level. The suspense measurer checks uncertainty and returns an estimated level of suspense when a sjuzhet is given as input. The interactions between these two components are controlled by the suspense creator, as described in Algorithm 1.

The algorithm initializes the content of the sjuzhet with the skeleton given from skeleton builder. In order to find supplemental story event contributing to suspense arousal, we define the term *potential suspense* that refers to the amount of each event's contribution to the suspense level increase (Heuristic Function 3). Our system selects e_s, the action with the greatest potential suspense from the set of satellite events S. If the potential suspense of e_s is lower than a predefined threshold, then the program returns S_T as the selected story events and exits. Otherwise, the system chooses an action e_k with the lowest importance in S_T and computes its potential suspense value $h(e_k,F)$. If the value is lower than the potential suspense of the new event ($h(e_s,F)$), the system builds a new partial plan $NewZ$ that replaces the event e_k with the event e_s from the set S_T.

Algorithm 1: The Suspense Creator procedure

input : F as a fabula plan containing a set of steps, a set of causal links, and a set of temporal constraints;

t as the step where the reader's suspense is measured;

S_K as the portion of the skeleton preceding t;

W as the set of importance weight for the events in F;

G as the protagonist's goal;

R as the reader's resource bound;

P as a planning algorithm;

L as the plan library as the reader's knowledge;

output: S_T as the content selected for suspense

$S_T \leftarrow S_K$;

$S \leftarrow$ (events in the portion of F preceding t) – S_K;

while S *is not empty* **do**

  $e_s \leftarrow$ an action in S with the highest $h(e_s, F)$;

  if $h(e_s, F) < threshold$ **then**

    Return S_T;

  end

  $S \leftarrow S - e_s$;

  $e_k \leftarrow$ an action in S_T with the lowest importance in W;

  if $h(e_s, F) > h(e_k, F)$ **then**

    $NewZ \leftarrow BuildPP(F, S_T - e_k + e_s)$;

    $OldZ \leftarrow BuildPP(F, S_T)$;

    if $NewZ$ *passes Uncertainty Checking* **then**

      if $sl(G, NewZ, L, P, R) > sl(G, OldZ, L, P, R)$ **then**

        $S_T \leftarrow S_T - e_k + e_s$;

      end

    end

  end

end

The function $BuildPP$ returns a partial plan structure which consists of plan steps and their causal links and temporal orderings coherent with F. Then the system queries the suspense measurer for the suspense levels of the new sjuzhet $NewZ$ and of the current sjuzhet $OldZ$ (computed by Heuristic Function 2). If the suspense level from $NewZ$ is higher than that of $OldZ$, the system updates the current S_T with the events in $NewZ$. This process continues until there is no candidate is found. When it terminates, the system specifies the content of the output sjuzhet as S_T.

A modification to the algorithm can create a story to elicit a low suspense level from the reader. The algorithm selects an action e_s with the lowest potential suspense and replaces the action e_k in S_T with the highest importance value if the current suspense level is lowered by replacing e_k with e_s.

The following sections detail the two subcomponents of the suspense creator. When a sjuzhet is given, the suspense measurer checks uncertainty condition and estimates suspense level based on Heuristic Function 2. The structure organizer constructs a sjuzhet by augmenting story events that are likely to increase suspense based on Heuristic Function 3, which makes use of the values calculated by Heuristic Function 4.

1) The Suspense Measurer: As discussed in Section II, one critical condition for a reader to feel suspense is to keep her uncertain about the outcome of a significant event. When the reader is certain about the negative outcome, she may feel disappointment or sadness rather than suspense [15]. To meet the uncertainty condition of suspense, the suspense measurer first checks if the reader model would be uncertain about the goal state using the planning space. In logical terms, an agent is uncertain about a proposition when the agent makes inferences, leading to the possibilities of being true and false [49]. The planning space represents the reader's reasoning and an inference corresponds to a path from the root node to a terminal node in the planning space. Therefore the reader model returns certainty when the planning space contains either only complete plans (absolute success) or only failed plans (absolute failure). Therefore, the reader model returns uncertainty about the goal state when the planning space has both successful plans and failed plans. The reader model also returns uncertainty when the planning space exceeds the searching limit, based on the rationale that successful plans can be found if unlimited resources are given.

When the uncertainty is ensured, the suspense measurer estimates the reader's suspense level, following the narrative comprehension model by Gerrig and Bernardo [16]. The reader's suspense increases as the number of options for the protagonist's successful outcome(s) decreases. Therefore, Heuristic Function 2 computes the reader's suspense level as the inverse of the number of planned solutions that achieve the protagonist's goal using her reasoning algorithm and her plan library within her reasoning limit. The function sets a minimum level of suspense when no usable solutions are found in her plan space, as is supported by psychological research.

2) Heuristic Function 2 (Level of Suspense): $sl(G, Z, L, P, R)$ returns the level of suspense when $success(G, Z, L, P, R) > 0$. Otherwise, it returns 0.

$$sl(G, Z, L, P, R) = \frac{1}{success(G, Z, L, P, R)}$$

where G is a set of literals representing the goal of a narrative's protagonist, Z is a partial plan, L is a plan library, P is a planning algorithm, R is an integer representing a reasoning bound, and $success(G, Z, L, P, R)$ returns the number of paths to make G true with given Z and R.

3) The Structure Organizer: The structure organizer selects a set of additional events α for suspense arousal, based on a heuristic function that examines the syntactical properties of a plan structure. The function is formalized based on the following rules assuming tha the reader is not informed of the protagonist's success or failure yet:

a) presenting a goal-threatening action whose effects negate the protagonist's goal/plan will increase the reader's suspense;
b) presenting a goal-supporting action whose effects unify with the protagonist's goal/plan will decrease the reader's suspense.

According to these rules, the elements of α would be composed of a set of goal-threatening actions to invoke a higher suspense. Note that the terms goal-threatening actions and goal-supporting actions are used unconventionally in this work. To determine whether an event is goal-threatening or not, we define the term *potential suspense*. Heuristic Function 3 computes the *potential suspense for an action* by summing up the *potential suspense* of its effects. An event is classified as a goal-threatening action if its potential suspense is greater than a predefined threshold. Conversely, an event is labeled as a goal-supporting action if its potential suspense is less than a predefined value.

4) Heuristic Function 3 (Potential Suspense of an Action): $h(a,p)$ returns the summation of $ps(e,a,p)$ where $ps(e,a,p)$ is the potential suspense of an effect e of an action a in plan p.

$$h(a,p) = \Sigma_{e \in E} PS(e,a,p)$$

where E is the set of a's effects, $ps(e,a,p)$ is the potential suspense of an effect e of an action a in plan p.

In computing the potential suspense of an action's effect, we consider all of the action's possible causal relationships to accomplishing the protagonist's goal from the reader's point of view. As an illustration, Fig. 4 shows that action *S5:Steal* has the effect $\neg(has\ Dad\ toy)$ —the negation of action *S2:Put* 's precondition *(has Dad toy)* —which can possibly interfere the action *S2:Put* with its execution from the reader's point of view until the successful outcome is disclosed. We call this type of temporary threats which are resolved later in the story as threatening-links, referring to an action's effect negating another step's precondition in the plan. In contrast, the suspense creator establishes a supporting-link when the effect of an action unifies with a precondition of another action in the plan. In the same figure, 4 *S1:Trade* has the effect *(has Dad toy)* that contains a supporting-link that unifies with the same condition of *S2:Put*. At the point where the action *S1:Trade* is shown the audience may expect that the goal of *getting Ben the toy* will be accomplished easily, although the later action *S5:Steal* will negate the *(has Dad toy)* condition. As illustrated, one effect can have multiple threatening-links and supporting-links in a single plan. Therefore, *potential suspense of an effect* is computed as the accumulation of all e's supporting-links weighted by w_s subtracted from the summation of all e's threatening-links weighted by w_t (Heuristic Function 4).

5) Heuristic Function 4 (Potential Suspense of an Effect): $PS(e,a,p)$ returns a value of an effect e of an action a in plan p as

$$\Sigma_{t \in T(e)} \frac{w_t}{DistEffect(a,p)} - \Sigma_{s \in S(e)} \frac{w_s}{DistEffect(a,p)}$$

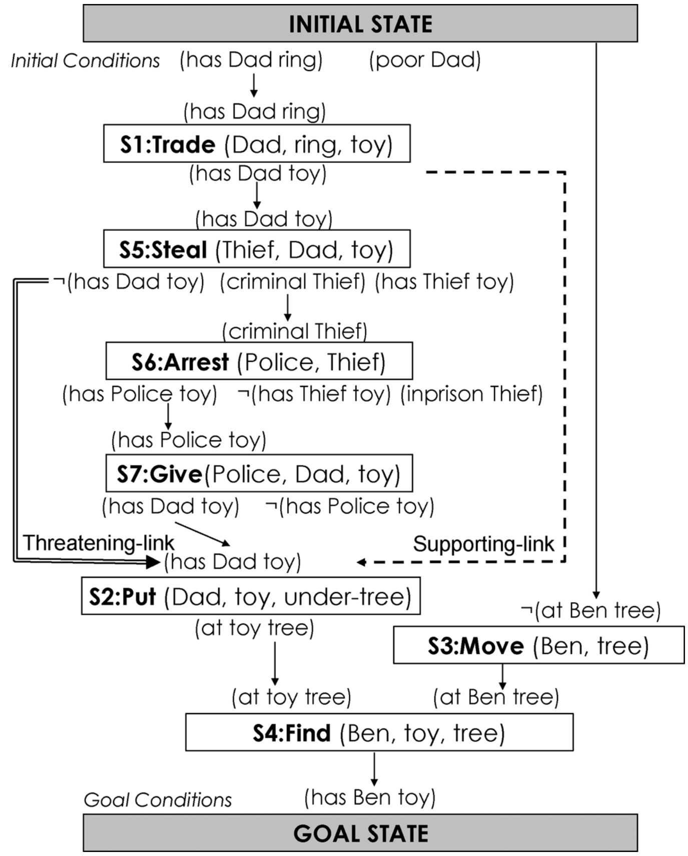

Fig. 4. Plan structure extended from Fig. 1: $S5$, $S6$, and $S7$ are added to illustrate threatening-links and supporting-links. These actions describe that the dad is stolen his toy by a thief ($S5$), the police arrests the thief ($S6$), and the police returns the toy to the dad ($S7$). A box represents an action, with its preconditions on the left and effects on the right. Solid arrows denote causal links. Dashed arrows indicate threatening-links in which the effect at the starting point negates the precondition of the action at the end point. Double-lined arrows represent supporting-links where a step's effect unifies with another step's precondition.

where $T(e)$ returns all the threatening-links of an effect e, $S(e)$ returns all the supporting-links of e, w_t and w_s are coefficients, and $DistEffect(a,p)$ returns a value associated with the causal distance between step a and the goal step of plan p. All scaling factors in the equation are constrained to be nonnegative real numbers.

V. Evaluation

This section describes several experiments that we carried out to evaluate Suspenser. An initial pilot study, described in the first section, aimed to test a partial implementation of Suspenser and the experimental methodology. The second section describes the main study we conducted.

A. A Pilot Study

We carried out a pilot study to evaluate the effectiveness of story generation for suspense produced by a partial implementation of Suspenser —the skeleton builder module (i.e., Heuristic Function 1) and the additional event selection by Heuristic Function 3. The reader model was not tested in this study; therefore, this pilot study did not evaluated the effectiveness of Heuristic Function 2 that measures the suspense level. The hypothesis for this study was to test if there was any association

TABLE I
EXPERIMENTAL VALUES FOR WEIGHTING CONSTANTS

Constant	Description	Value
k_i	incoming causal links	1.0
k_j	incoming causal links from the initial state	0.3
k_o	outgoing causal links	5.0
k_c	category	2.5
w_t	threatening-link	7.0
w_s	supporting-link	2.0

between the story generator type (independent variable) and the suspense level of the stories (response variable). To test this hypothesis, we compared the suspense levels among the stories produced by: a) Suspenser to elicit high suspense; b) a human author; and c) a baseline that is designed to elicit low suspense in the reader.

1) Configuring the Experimental System: The values of the scaling factors for our heuristic functions are shown in Table I. The values of the constant factors were determined empirically from some informal experiments. In our study the value for k_i and k_o, were initialized as 1 and 5, respectively, informed by the CPI model that generates concise instructions [46]. k_j and k_c were adjusted as 0.3 and 2.5, respectively. The setting of these coefficients placed more importance on the value of outgoing causal links than the incoming causal links, which is consistent with the views discussed in the cognitive research [47], [50]. The causal-chain value of an event was assigned 2.0 for *motivated-act* ; the value of an event of other act types was assigned 0.0. The value for the threatening-link coefficient (w_t) was assigned greater than that of the supporting-link coefficient (w_s). These values were chosen through informal experiments and pilot studies such as [51] for summarization and [52] for event selection. The experimenter manually tested numerous combinations of values and selected the one that produced the most suspenseful story based on her subjective judgements. The optimal combination could be found by examining all possible combinations and having multiple judges. The function that reflects an event's distance effect, $DistEffect(a, p)$ in Heuristic Function 1 returned 1 and $DistEffect(a, p)$ in Heuristic Function 4 returned the distance from an action to the goal (i.e., the minimum number of causal links that relate an action to the goal in a plan).

2) Materials: We ran Crossbow to plan three fabulas. The resulting plans consisted of 18 to 20 partially ordered steps which were manually linearized. Each plan was realized as text using a simple template-matching technique that mapped one plan step into a single sentence. We prepared a total of nine sjuzhets by generating three sjuzhets for each fabula: two sjuzhets by Suspenser and one sjuzhet by a human author. To obtain human generated stories, we recruited one Master's student majoring in English at North Carolina State University, a freelance writer who had previously had short stories published in a local newspaper. She was presented with the instruction sheet followed by the three fabulas and their corresponding measurement point. She then was asked to select a series of sentences for each fabula to elicit high suspense from the reader at the specified point in the story. We did not constrain the number of sentences that she selected.

TABLE II
MEANS AND STANDARD DEVIATIONS OF SUSPENSE RATINGS

Generator	Story	N	Mean	SD
Suspenser	Fabula A	9	2.22	0.67
	Fabula C	8	3.25	1.28
	Total	17	2.71	1.05
Human	Fabula A	8	2.63	1.19
	Fabula C	8	2.75	0.04
	Total	16	2.69	1.08
Control	Fabula A	8	2.25	1.16
	Fabula C	9	3.00	1.58
	Total	17	2.65	1.41

3) Procedure: We utilized a repeated measured between group design and assigned the subjects randomly to one of nine subject groups. These groups were arranged according to a 3×3 Latin Square design to counterbalance the interference from different orderings of stories. Each subject individually participated in the study by accessing a web page that presented a sequence of three stories. Each story was divided into two parts: one containing the text describing the story's background and the portion preceding the measurement point in the story, and one containing a paragraph describing the portion of the story after the measurement point. After reading the first part on a web page, the subject was asked to click the button “NEXT PAGE” to proceed to the next screen in which he was asked to answer his suspense level on a seven-point Likert scale of [1, 7] where 1 denotes *no suspense* and 7 denotes *extremely suspenseful*. On the completion of responding to the question, a button click led him to the next page that described the second part of the story. The subject was able to leave the survey by closing the survey web page anytime they wanted.

4) Results: A total of 25 undergraduate students ranging in age from 20 to 29 years old participated in this study. There were 23 males and 2 females, all recruited from a Computer Science undergraduate course at the North Carolina State University. They were given extra credit in exchange of participating in this study, and were presented an alternative option. The collected data contained 75 responses from 25 subjects, 25 responses for each fabula. However, due to an error in reproduction of the writer's selection, the responses for sjuzhets created from Fabula B were excluded from the analysis. As a result, 50 observations were used in this analysis. To detect a significant difference between story generators and fabulas, we performed a two-way ANOVA to the collected data using SAS version 9.1.3 SP4. As shown in Tables II and III, the data indicated that the story generator had no effect on the suspense level ($\mathrm{F}(2, 44) = 0.01$, $\mathrm{p} = 0.99$). The story generators showed uniform performance across the two stories. There was no statistical difference in suspense ratings for fabula ($\mathrm{F}(1, 44) = 3.64$, $\mathrm{p} = 0.06$). No interaction effect was found between the fabula type and the story generator type ($\mathrm{F}(2, 44) = 0.64$, p value $= 0.53$).

5) Discussion: The stories generated by Suspenser obtained the highest mean ratings (Table II shows). However, no sta-

TABLE III
ANOVA Summary Table for Suspense

Source	DF	SS	MS	F Value	Pr > F
Fabula	1	5.12	5.12	3.64	0.06
Generator	2	0.03	0.02	0.01	0.99
Fabula*Generator	2	1.80	0.90	0.64	0.53
Error	44	61.93	1.41		

tistical significant difference in suspense has been found in the story generator type, as the ANOVA analysis indicates (Table III). We conjecture that this result may be caused by several factors described below. First, the number of subjects was too small to detect a difference, particularly with short, similar stories. Since the skeleton serves as a base story when creating the stories intended for high suspense and low suspense, these stories share more than 50% of the total number of story sentences. Therefore, we needed to find another baseline stories that are distinct from the other stories. Another reason relates to the way of presenting the story to the participants. A whole story was presented to the subjects at once on a single web page, which normally took them a very short time to read. Thus, the subjects did not have enough time to prepare themselves to anticipate a next story event.

From the lessons learned from the result of this pilot study, we made some changes to the design of the main experiments. First, the survey interface was modified to give the subject time to build expectation about the next story segment. One web page shows only a single story event and the subject was required to click a button to proceed to read the next event. Second, the subject's suspense was measured on a five-level scale.

B. The Main Experiment

This section describes the experiment that we carried out to evaluate the effectiveness of stories that a complete implementation of Suspenser produces in terms of suspense. The hypothesis for our study was to test if there was any association between the story generator type (the independent variable) and the suspense level (the dependent variable). To test this hypothesis, we compared the suspense levels reported by subjects as they read stories produced by: a) Suspenser; b) a human author intended to create high suspense; and c) the human author intended for low suspense as a control narrative.

1) Configuring the Experimental System: Table I shows the values of the scaling factors for our heuristic functions for this study. The function that reflects an event's distance effect, $DistEffect(a, p)$ in Heuristic Function 1 and Heuristic Function 4 returned $d \times (d + 1)$ where d denotes the distance from an action to the goal. The value of *threshold* in the algorithm 1 was assigned 0.07. The reader's knowledge was assumed identical to the system's plan libraries that were used to create the input fabulas. The reader model's plan ranking function preferred short plans with fewer flaws. Its reasoning resource limit was set to a search limit of 500 nodes.

2) Materials: For the main experiment, the three fabulas in the pilot study were reused. We prepared a total of nine sjuzhets by generating three sjuzhets for each fabula:one sjuzhet by Suspenser and two sjuzhets by a human author. The human writer who chose the story events for the pilot study was asked to select two series of sentences for each fabulas: one to arouse high suspense and the other to elicit low suspense form the reader when his suspense level would be measured at a given point in the story. Due to the short length of stories generated by the system, a single measurement point was selected close to the ending but where a clear outcome is not revealed yet. For instance, the measurement point for Fabula C was after reading 17th sentence (see Appendix A). When a series of story events preceding the outcome hints the outcome, the measurement point was selected before the those preceding events occur. The measurement point for each story was presented to the human author before her selection. Her selection was not constrained; as a result, her two versions of a fabula differed in length within a margin of 2 sentences. For fair comparisons, the length of the sjuzhet (N) generated by the system was set to the number of the story events in the corresponding human author's selection for high suspense up to the point where the reader's suspense was measured. The human author's selection for eliciting low suspense for each fabula was used as a control narrative. The control narratives differed significantly in content from the other stories. While the stories created by Suspenser and those by the human author for arousing high suspense share 50%-80% of the total number of story sentences, the control stories overlap with the other stories about 20%-30% of the total number of story sentences. One sample of the fabulas and its three sjuzhets used in this study are shown in Appendix A.

3) Procedure: The study utilized a repeated measured between group design: subjects were randomly assigned to one of nine subject groups that were designed based on a 3×3 Latin Square. Each subject individually participated in the study by accessing a web page. He was presented with three stories and was asked to rate his suspense level for each story at a given point. Each story was presented to the subject sentence by sentence; one page contained only one sentence and a button click led to the next page. After reading the portion of the text preceding the measurement point, the subject was asked to describe his suspense level on a five-point scale basis ranging from *no suspense* to *extremely suspenseful*. After responding to the question, he was presented with the second part of the story sentence by sentence, followed by a page asking generic questions about story coherence and enjoyment on a five-point scale ranging from 1 meaning *not at all* to 5 meaning *strongly agree*.

4) Data Collection: A total of 98 unpaid subjects voluntarily took part in the experiment, ranging in age from 20 to more than 50 years old (42 males, 57 females, and one no response): 72 recruited from NCSU communities including recently graduated under/graduate students across different departments and 26 from internet female technical communities (e.g., Systers.org). All subjects were native-speakers of English. The collected data contained 294 responses from 98 subjects.

5) Results: Table IV shows that the suspense ratings for the stories selected by Suspenser are as high as the ratings for the stories chosen by the human author. To detect a significant difference between three story generators, we performed a two-way ANOVA on the collected data using SAS version 9.1.3 SP4. In this analysis, two main effects were examined: the story generator type and the fabula type. Each type has three levels.

TABLE IV
MEANS AND STANDARD DEVIATIONS FOR SUSPENSE IN EACH STORY AND STORY GENERATOR

Generator	Story	N	Mean	SD
Suspenser	Fabula A	32	2.44	0.91
	Fabula B	33	2.73	1.13
	Fabula C	33	2.94	1.09
	Total	98	2.70	1.06
Human	Fabula A	33	2.67	0.89
	Fabula B	32	2.66	1.10
	Fabula C	33	2.76	1.17
	Total	98	2.69	1.05
Control	Fabula A	33	2.30	1.10
	Fabula B	33	2.39	1.14
	Fabula C	32	2.25	0.95
	Total	98	2.32	1.06

TABLE V
ANOVA SUMMARY TABLE FOR SUSPENSE

Source	DF	SS	MS	F Value	Pr > F
Fabula	2	1.71	0.86	0.76	0.47
Generator	2	9.57	4.79	4.27	0.02
Fabula*Generator	4	2.95	0.74	0.66	0.62
Error	285	319.76	1.12		

TABLE VI
ONE-TAILED t-TEST ANALYSIS SHOWING PAIR-WISE COMPARISON OF MEANS FOR SUSPENSE. COMPARISONS SIGNIFICANT AT THE 0.01 LEVEL ARE INDICATED BY **

Generator Comparisons	t Value	Pr > | t |
Suspenser vs. Human	0.07	0.473
Suspenser vs. Control	2.56	0.006**
Human vs. Control	2.50	0.007**

As shown in Table V, the data indicated that the story generator type had an effect on the suspense level ($\mathrm{F}(2, 285) = 4.27$, p value $= 0.02$). The result also shows that the fabula type had no effect on suspense. No interaction effect was found between the fabula type and the story generator type ($\mathrm{F}(4, 285) = 0.66$, p value $= 0.62$). Despite the short sample stories, the subjects rated their experience of suspense as *moderate* (mean $=$ $2.57/5.0$, SD $= 1.06$) on a five-point Likert scale of [1, 5] where 1 means *no suspense* and 5 means *extremely suspenseful*. A series of standard one-tailed t-tests were used to compare the performance of the three story generators. The results in Table VI indicate that the participants were more likely to rate the stories produced by Suspenser and the human author more suspenseful than the control versions with a 99% of confidence level ($t = 2.56$, p value $= 0.006$ for Suspenser versus Control; $t = 2.50$, p value $= 0.007$ for Human versus Control). The effect sizes were small-to-moderate (Cohen's d $= 0.36$, power $= 0.807$ with a 0.05 of significance level for Suspenser versus Control; Cohen's d $= 0.35$, power $= 0.787$ with a 0.05 of significance level for Human versus Control).

To test if the text quality affected the reader's story comprehension, the subjects' responses to story coherency were also analyzed. The data suggest that the text quality was good enough for the subjects to understand the stories. The participants evaluated the given stories as moderately coherent (mean $= 2.36/5.0$, SD $= 0.97$) on the five Likert scale of [1, 5] where 1 means *not coherent* and 5 means *strongly coherent*. We further investigated the reader's suspense for only those stories that received valid coherency ratings. To this end, we eliminated the reports that contain *not coherent* or *no response* on the coherency question. This preprocessing step eliminated 70 responses from the initial data, resulting in 224 responses. For the data analysis the R version 2.14.2 software package was used [53]. As Table VII shows, the stories created by the human writer received the highest ratings in suspense (mean $= 3.03$). On the other hand, Suspenser's stories are superior to human author's stories in terms of interestingness (mean $= 3.42$) and coherence (mean $= 2.82$). A two-way ANOVA on the data (Table VIII) indicated that both the fabula and story generator types had effects on the suspense level (p value $= 0.02$). No interaction effect was found between the fabula type and the story generator type. Pearson product-moment correlation analysis indicates high positive correlations among suspense, interestingness, and coherence ($\mathrm{p} < 0.01$). The results of one-tailed t-tests (in Table IX) indicate that the stories produced by the system and the human author were rated as more suspenseful than the control stories with a 95% confidence interval and a 99% confidence interval, respectively ($t = 2.10$, p value $= 0.018$ for Suspenser versus Control; $t = 2.62$, p value $= 0.005$ for Human vs. Control). The effect size

TABLE VII
MEANS AND STANDARD DEVIATIONS FOR SUSPENSE, INTERESTINGNESS, AND COHERENCE RATINGS FOR EACH STORY GENERATOR TYPE IN COHERENT NARRATIVE EXPERIENCE ONLY

Generator (No. of responses)	Suspense Mean (SD)	Interestingness Mean (SD)	Coherence Mean (SD)
Suspenser (78)	2.96 (0.99)	3.42 (0.96)	2.82 (0.83)
Human (76)	3.03 (0.89)	3.13 (0.99)	2.75 (0.71)
Control (70)	2.63 (0.94)	2.63 (2.83)	2.61 (0.67)

TABLE VIII
ANOVA SUMMARY TABLE FOR SUSPENSE IN COHERENT NARRATIVE EXPERIENCE ONLY

Source	DF	SS	MS	F Value	Pr > F
Fabula	2	6.96	3.48	4.10	0.02
Generator	2	7.01	3.50	4.13	0.02
Fabula*Generator	4	5.17	1.29	1.52	0.20
Error	215	182.61	0.85		

TABLE IX
ONE TAILED t-TEST ANALYSIS SHOWING COMPARISON OF MEANS FOR SUSPENSE. COMPARISONS SIGNIFICANT AT THE 0.01 LEVEL ARE INDICATED BY ** AND COMPARISONS SIGNIFICANT AT THE 0.05 LEVEL ARE INDICATED BY *

Generator Comparisons	t Value	Pr > | t |
Suspenser vs. Human	0.43	0.335
Suspenser vs. Control	2.10	0.018*
Human vs. Control	2.62	0.005**

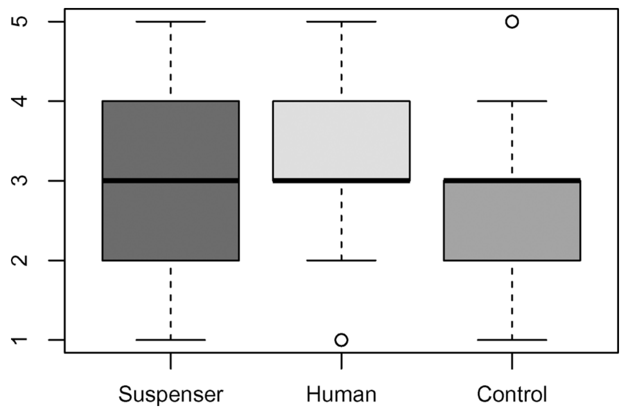

Fig. 5. Box plot graphs of suspense ratings for coherent stories only. The X axis represents the story generator types and the Y axis represents suspense ratings. Each box accounts for the 50% of the population and the thick horizontal bar in the box represents the median value. The top of the box denotes the upper quartile and its bottom denotes the lower quartile. Those ratings in the top 25% of the data are shown by the top whisker and those in the low 25% are shown in the bottom whisker. The dots represent outliners.

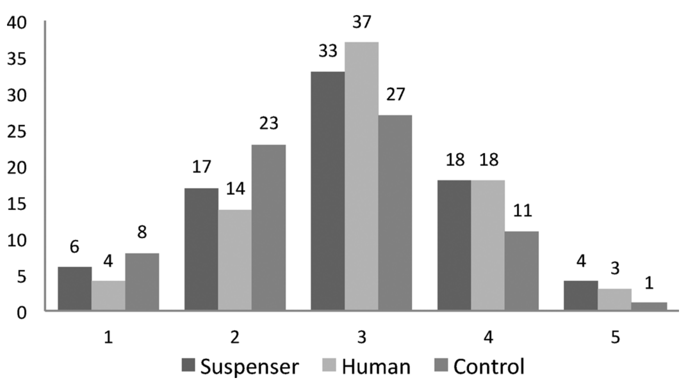

Fig. 6. Histograms of suspense ratings for coherent story responses only. The X axis represents the suspense ratings and the Y axis represents the count of responses.

was small-to-moderate for Suspenser versus Control (Cohen's $d = 0.34$, power $= 0.64$ with a 0.05 significance level) and moderate for Human versus Control (Cohen's $d = 0.44$, power $= 0.83$ with a 0.05 significance level). The boxplot diagrams in Fig. 5 depict that the suspense ratings for stories by human author are skewed to the high ratings and the ratings of control stories are skewed to the low ratings. The histogram in Fig. 6 shows the counts of reports for each suspense ratings grouped by the story generator.

6) Discussion: The data show that the story generators had an influence on the amount of suspense that the subjects felt. In particular, the stories produced by Suspenser created stories comparable in suspense to those produced by human authors. The one-tailed t-test results also show that the difference between the suspense levels felt by the subjects from Suspenser's stories and the control stories was significant (a 99% confidence interval for all stories and a 95% confidence interval for coherently rated stories only).

For the coherently rated story data set, the human author's stories received the highest suspense ratings, resulting in the largest effect size when compared with the stories in the control group (Cohen's $d = 0.44$). It is also noted that the distributions of suspense ratings for Suspenser's stories and the human author's stories were different. Compared with the relatively even distribution in suspense ratings for Suspenser, the ratings for the author's stories were inclined to the high-rating side. This may suggest that there can be qualitative differences between the stories generated by Suspenser and those generated by the human author. On the other hand, the stories generated by the system were rated higher in interestingness and coherence than those generated by the human author. This result could be linked to the fact that the human author was asked to select story events for the effect of suspense only, whereas Suspenser considers coherence as well by building the skeleton from the story's core story events. Therefore, further discussions regarding interestingness and coherence are beyond the scope of this paper that focuses on suspense only.

VI. Conclusion

Narrative generation by computers has been actively researched for two decades with special attention to games. Although a number of approaches have shown promise in their ability to generate narrative, there has been little research on creating stories that prompt an intended emotion. This paper presents a computational model of generating stories for suspense, exploring the concept that a reader's suspense level is affected by the number of solutions available to the problems faced by a narrative's protagonists [15], [16], [22], [25], [26]. When given a formal characterization of a story world, this model elaborates a story content that can manipulate reader suspense at a specific point in its telling. Our approach gauges the suspense level that a reader would feel by modeling the reader's narrative comprehension process using a planning technique. The system takes as input a partial plan corresponding to the portion of a story that has been conveyed so far and computes the reader's anticipated suspense level based on the inverse of the number of solutions to the protagonist's goals that can be found in the space of complete plans she can consider within her reasoning resources. To generate a partial plan that maximizes the reader's suspense, the system takes a plan as input and selects a set of core events that have high causal connectivity and that also play an important role in the story. The partial plan then is supplemented by harmful actions that intensify the reader's suspense level. The model has been implemented and formally evaluated. The quantitative analyses of the data obtained from the experiments have shown this system to be successful in selecting content that elicits high suspense. In particular, the data show that, in the context of our experiments, this model was as effective as a human author in generating suspenseful stories. To our knowledge, Suspenser is the first system that aims to generate suspenseful stories by modeling a storyteller who selects relevant story elements based on the reader's reasoning process. We believe that this work will benefit the AI, game, and affective computing communities.

The most important limitation of the current study is the heuristic function that approximates the suspense level by the number of solutions for the protagonist's goal only, without examining the likelihood of their successful executions. Despite this limitation, Suspenser has shown successful results in selecting a suspenseful story from a given story when compared with a human author's performance. We believe that this result was partly due to its subcomponents: the skeleton builder and the structure organizer. The core events identified by the skeleton builder may ensure that the final sjuzhet contains the suspensefulness that the input fabula is able to elicit. In addition, the structure organizer selects candidate events for inclusion in the sjuzhet as the events whose effects are negating another event's preconditions. Therefore, we conclude that the limitation of the heuristic function in the suspense measure component could be compensated by the other components in the system. A second limitation concerns the evaluation that uses a single human author. We recruited a person who is a professional writer, expecting that she would represent the expert in narrative generation. However, the task of selecting sjuzhet from a predefined story events may leave little room to express her expertise. In the current study, the writer selected 31 out of 57 events of all the fabulas, accounting for 54%. It is expected that the writer will exhibit a better performance when a wider selection is offered. A further evaluation can be conducted by having fabulas consisting of a large number of events and by recruiting multiple human authors.

Several aspects of the current system will be investigated and extended in future work. First, we will refine the function used to measure the suspense level in a story. For instance, taking into account the difficulty of achieving a plan (e.g., size, readiness of executing its actions) in the human reader's reasoning process can be simulated by employing a probabilistic planning technique. Second, the current approach uses a plan library that is used for generating the input fabula plan as the reader's knowledge. It will be interesting and challenging to investigate the problem of knowledge discrepancies in storytelling via the use of different plan libraries.

Appendix

Fabula C. The Parts That Were Shown After Suspense Measurement Were Italicized

Background: Sykes is the owner of the Hollywood Theater, which was once prosperous but has now become dilapidated and is in need of major renovations. Sykes has accrued a sizable gambling debt, and with his theater in shambles, he has no means with which to pay it back. He is constantly threatened by his crooked debtors. Janet is a famous actress with dreams of winning an Oscar, an acting award. She is jealous of the actress Agatha, who is her contender for the Oscar this year and also is well known for her active involvement in charity. Janet knows a number of scoundrels including a guy named Kent, a bomb dealer, and the theater owner Sykes. Agatha is in love with Bill, who serves as a lieutenant in the Los Angeles Police Department's Serious Crime squad. Janet knows that Agatha is planning to go to the Charity Bazaar for the Poor to be held in Hollywood Theater. To ensure that she will win the Oscar, Janet plans to kill Agatha during the charity event.

The Input Fabula Story: [1] Janet convinces Sykes to participate in her plan to kill Agatha by convincing him that if he participates, he will be able pay off his gambling debts. [2] Janet and Sykes plan to burn down Sykes' theater to get the insurance money and kill Agatha during the charity bazaar. [3] Sykes borrows some money from the bank by mortgaging his theater. [4] Sykes buys insurance to cover his loss in case of a fire. [5] Janet gives Kent's contact information to Sykes and informs him of Kent's expertise with firebombs. [6] Kent takes a bomb to the Hollywood Theater and meets with Sykes. [7] Sykes purchases the firebomb. [8] Sykes installs the firebomb. [9] The lieutenant, Bill, issues a warrant permitting the arrest of Kent for his illegal weapons dealing. [10] Bill arrests Kent. [11] Bill coaxes Kent to give information in exchange for releasing him. [12] Kent informs Bill that Sykes is planning to firebomb his own theater during the charity event. [13] Bill releases Kent for his cooperation. [14] Agatha goes to the theater for the charity event. [15] Sykes sets the timer of the firebomb to explode during the charity event. [16] Sykes switches on the firebomb. [17] Bill searches for the firebomb in the theater. *[18] Bill defuses the firebomb. [19] Agatha participates in the charity event.*

Sjuzhet 1: Storywriter's Selection Intended for High Suspense: Janet and Sykes plan to burn down Sykes' theater to get the insurance money and kill Agatha during the charity bazaar. Janet gives Kent's contact information to Sykes and informs him of Kent's expertise with firebombs. Kent takes a bomb to the Hollywood Theater and meets with Sykes. Sykes purchases the firebomb. Sykes installs the firebomb. Kent informs Bill that Sykes is planning to firebomb his own theater during the charity event. Agatha goes to the theater for the charity event. Sykes sets the timer of the fire-bomb to explode during the charity event. Sykes switches on the firebomb. Bill searches for the firebomb in the theater. *Bill defuses the firebomb.*

Sjuzhet 2: Suspenser's Selection Intended for High Suspense: Kent takes a bomb to the Hollywood Theater and meets with Sykes. Sykes purchases the firebomb. Sykes installs the firebomb. Bill arrests Kent. Kent informs Bill that Sykes is planning to firebomb his own theater during the charity event. Bill releases Kent for his cooperation. Agatha goes to the theater for the charity event. Sykes sets the timer of the firebomb to explode during the charity event. Sykes switches on the firebomb. Bill searches for the firebomb in the theater. *Bill defuses the firebomb. Agatha participates in the charity event.*

Sjuzhet 3: Storywriter's Selection for Low Suspense as a Baseline Story: Janet convinces Sykes to participate in her plan to kill Agatha by convincing him that if he participates, he will be able pay off his gambling debts. Sykes borrows some money from the bank by mortgaging his theater. Sykes buys insurance to cover his loss in case of a fire. Janet gives Kent's contact information to Sykes and informs him of Kent's expertise with firebombs. Kent takes a bomb to the Hollywood Theater and meets with Sykes. Sykes purchases the firebomb. The lieutenant, Bill, issues a warrant permitting the arrest of Kent for his illegal weapons dealing. Bill coaxes Kent to give information in exchange for releasing him. Bill releases Kent for his cooperation. *Agatha participates in the charity event.*

References

[1] *Heavy Rain*. Quantic Dream, 2010.
[2] *L.A. Noire*. Team Bondi, 2011.
[3] *BioShock Infinite*. Irrational Games, 2013.
[4] *Fallout: New Vegas*. Obsidian Entertainment, 2010.

[5] M. Cavazza, F. Charles, and S. J. Mead, "Character-based interactive storytelling," *IEEE Intell.t Syst.*, vol. 17, pp. 17–24, Jul. 2002.
[6] M. Mateas and A. Stern, "Façade: An experiment in building a fully-realized interactive drama," in *Proc. Game Develop. Conf.*, 2003.
[7] B. Mott, S. Lee, and J. Lester, "Probabilistic goal recognition in interactive narrative environments," in *Proc. AAAI*, 2006.
[8] R. W. Hill Jr., A. W. Hill, J. Gratch, S. Marsella, J. Rickel, W. Swartout, and D. Traum, "Virtual humans in the mission rehearsal exercise system," *KI Embodied Conversat. Agents*, vol. 17, pp. 32–38, 2003.
[9] N. Szilas, "A new approach to interactive drama: From intelligent characters to an intelligent virtual narrator," in *Proc. AAAI Spring Symp. AI Interact. Entertain.*, 2001, pp. 72–76.
[10] W. F. Brewer and E. H. Lichtenstein, "Stories are to entertain: A structural-affect theory of stories," *J. Pragmat.*, vol. 6, no. 5-6, pp. 473–486, Dec. 1982.
[11] W. F. Brewer and K. Ohtsuka, "Story structure, characterization, just world organization, and reader affect in american and hungarian short stories," *Poetics*, vol. 17, no. 4-5, pp. 395–415, 1988.
[12] L. F. Alwitt, "Suspense and advertising responses," *J. Consumer Psychol.*, vol. 12, no. 1, pp. 35–49, 2002.
[13] A. Ortony, G. L. Clore, and A. Collins, *The Cognitive Structure of Emotions*. Cambridge, U.K.: Cambridge Univ. Press, 1988.
[14] P. Vorderer, H. J. Wulff, and F. Mike, "Toward a psychological theory of suspense," in *Suspense: Conceptualizations, Theoretical Analyses, and Empirical Explorations*. Mahwah, NJ, USA: Lawrence Erlbaum Associates, 1996.
[15] D. Zillmann, "The psychology of suspense in dramatic exposition," in *Suspense: Conceptualizations, Theoretical Analyses, and Empirical Explorations*, P. Vorderer, H. J. Wulff, and M. Friedrichsen, Eds. Mahwah, NJ: Erlbaum, 1996, pp. 199–232.
[16] R. J. Gerrig and A. B. I. Bernardo, "Readers as problem-solvers in the experience of suspense," *Poetics*, vol. 22, no. 6, pp. 459–472, 1994.
[17] G. Genette, *Narrative Discourse*. Oxford, U.K.: Basil Blackwell, 1980.
[18] M. Bal, *Narratology: Introduction to the Theory of Narrative*. Toronto, ON, Canada: University of Toronto Press, 1985.
[19] S. Rimmon-Kenan, *Narrative Fiction: Contemporary Poetics*, 2nd ed. London, U.K.: Routledge, 2002, New accents.
[20] M. J. Toolan, *Narrative: A Critical Linguistic Introduction*, 2nd ed. London, U.K.: Routledge, 2001, Interface.
[21] D. Zillmann, "The logic of suspense and mystery," in *Responding to the Screen: Reception and Reaction Processes*, Bryant and D. Zillmann, Eds. Hillsdale, NJ, USA: Lawrence Erlbaum, 1991, pp. 281–303.
[22] W. F. Brewer, "The nature of narrative suspense and the problem of rereading," in *Suspense: Conceptualizations, Theoretical Analyses, and Empirical Explorations*, P. Vorderer, H. J. Wulff, and M. Friedrichsen, Eds. Mahwah, NJ, USA: Lawrence Erlbaum Associates, 1996, pp. 107–127.
[23] R. J. Gerrig, "The resiliency of suspense," in *Suspense: Conceptualizations, Theoretical Analyses, and Empirical Explorations*, P. Vorderer, H. J. Wulff, and M. Friedrichsen, Eds. Mahwah, N.J., USA: Lawrence Erlbaum Associates, 1996, pp. 93–105.
[24] P. Comisky and J. Bryant, "Factors involved in generating suspense," *Human Commun. Res.*, vol. 9, no. 1, pp. 49–58, 1982.
[25] N. Carroll, "The paradox of suspense," in *Suspense: Conceptualizations, Theoretical Analyses, and Empirical Explorations*, P. Vorderer, H. J. Wulff, and M. Friedrichsen, Eds. Mahwah, NJ, USA: Lawrence Erlbaum Associates, 1996, pp. 71–92.
[26] M. de Wied, "The role of temporal expectancies in the production of film suspense," *Poetics*, vol. 23, no. 1-2, pp. 107–123, 1995.
[27] N. M. Sgouros, "Dynamic generation, management and resolution of interactive plots," *Artif. Intellig.*, vol. 107, pp. 29–62, Jan. 1999.
[28] M. Lebowitz, "Story-telling as planning and learning," *Poetics*, vol. 14, no. 6, 1985.
[29] B. Magerko and J. E. Laird, *Mediating the Tension Between Plot and Interaction*. San Jose, CA: , 2004, 2004 AAAI Workshop Series: Challenges in Game Artificial Intelligence.
[30] R. Aylett, "Narrative in virtual environments-towards emergent narrative," in *Proc. Narrative Intell. Symp.*, 1999.
[31] S. R. Turner, *The Creative Process: A Computer Model of Storytelling and Creativity*. Hillsdale, NJ, USA: Lawrence Erlbaum Assoc., 1994.
[32] M. O. Riedl and R. M. Young, "Narrative planning: Balancing plot and character," *J. Artif. Intell. Res.*, vol. 39, 2010.
[33] S. G. Ware and R. M. Young, "CPOCL: A narrative planner supporting conflict," in *Proc. AIIDE*, 2011.
[34] R. M. Young, M. E. Pollack, and J. D. Moore, "Decomposition and Causality in Partial Order Planning," in *Proc. 2nd Int. Conf. Artif. Intell. Plan. Syst. (AIPS)*, 1994, pp. 188–193.
[35] R. M. Young and J. D. Moore, "DPOCL: A principled approach to discourse planning," in *Proc. 7th Int. Workshop Natural Lang. Generat.*, 1994, pp. 13–20.
[36] D. Pizzi, F. Charles, J. luc Lugrin, and M. Cavazza, "Interactive storytelling with literary feelings," in *Proc. 2nd Int. Conf. Affect. Comput. Intell. Interaction (ACII2007)*, 2007.
[37] F. Peinado, M. Cavazza, and D. Pizzi, "Revisiting character-based affective storytelling under a narrative BDI framework," in *LNCS 5334: Interactive Storytelling*, U. Spierling and N. Szilas, Eds. Berlin, : Springer-Verlag, 2008, pp. 83–88.
[38] J. Platts, A. Blandford, and C. Huyck, "Awaiting the sensation of a short, sharp shock: Twist-centred story generation by transformation," in *Proc. AISB'02 Symp. AI Creativity Arts Sci.*, 2002.
[39] B.-C. Bae and R. M. Young, "A use of flashback and foreshadowing for surprise arousal in narrative using a plan-based approach," in *LNCS 5334: Interactive Storytelling*, U. Spierling and N. Szilas, Eds. Berlin, Germany: Springer-Verlag, 2008, pp. 156–167.
[40] C. B. Callaway and J. C. Lester, "Narrative prose generation," *Artif. Intell.*, vol. 139, pp. 213–252, Aug. 2002.
[41] A. Jhala and R. M. Young, "Cinematic visual discourse: Representation, generation, and evaluation," *IEEE Trans. Computat. Intell. AI in Games*, vol. 2, no. 2, pp. 69–81, Jun. 2010.
[42] R. Barthes and L. Duisit, "An introduction to the structural analysis of narrative," *New Literary History*, vol. 6, no. 2, pp. 237–272, 1975.
[43] S. B. Chatman, *Story and Discourse: Narrative Structure in Fiction and Film*. Ithaca, NY, USA: Cornell Univ. Press, 1980.
[44] M. J. Rattermann, L. Spector, J. Grafman, H. Levin, and H. Harward, "Partial and total-order planning: Evidence from normal and prefrontally damaged populations," *Cogn. Sci.*, vol. 25, no. 6, pp. 941–975, 2001.
[45] S. Kambhampati, C. A. Knoblock, and Q. Yang, "Planning as refinement search: A unified framework for evaluating design tradeoffs in partial-order planning," *Artif. Intell.*, vol. 76, no. 1-2, pp. 167–238, 1995.
[46] R. M. Young, "Using Grice's maxim of quantity to select the content of plan descriptions," *Artif. Intell.*, vol. 115, pp. 215–256, Dec. 1999.
[47] T. Trabasso and L. L. Sperry, "Causal relatedness and importance of story events," *J. Memory Lang.*, vol. 24, no. 5, pp. 595–611, 1985.
[48] C. L. Foss and G. H. Bower, "Understanding actions in relation to goals," in *Advances in Cognitive Science*, N. E. Sharkey, Ed. Chichester, U.K.: Ellis Horwood, 1986, vol. I, pp. 94–124.
[49] W. van der Hoek and A. Lomuscio, "A logic for ignorance," in *LNCS 2990: Declarative Agent Languages and Technologies*. Berlin, Germany: Springer-Verlag, 2004, pp. 97–108.
[50] A. C. Graesser, K. L. Lang, and R. M. Roberts, "Question answering in the context of stories," *J. Experiment. Psychol.: General*, vol. 120, no. 3, pp. 254–277, 1991.
[51] Y.-G. Cheong and R. M. Young, "A computational model of narrative generation for suspense," in *Proc. 21st AAAI 2006 Computat. Aesthetic Workshop*, 2006, pp. 1906–1907.
[52] Y.-G. Cheong, A. Jhala, B.-C. Bae, and R. M. Young, "Automatically generating summary visualizations from game logs," in *Proc. AIIDE*, 2008.
[53] R. D. C. Team, *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation, 2012, Statistical Computing.

Yun-Gyung Cheong (M'10) received the B.S. degree, in 1996, and the M.S. degree, in 1998, in information engineering from Sungkyunkwan University, South Korea. In 2007, she received the Ph.D. degree in computer science from North Carolina State University, Chapel Hill, NC, USA.

She worked as a Researcher at Samsung Advanced Institute of Technology, South Korea (2008–2010), and was a Postdoctoral Fellow of the Center for Computer Games Research, IT University of Copenhagen, in Denmark (2010–2014). She is currently a Research Professor at the College of ICE at Sungkyunkwan University. Her research interests include artificial intelligence with emphasis on its use in narrative planning, games, and data mining.

Dr. Cheong is a Member of the IEEE Games Technical Committee (GTC) of the Computational Intelligence Society.

R. Michael Young (M'03–SM'04) received the B.S. degree in computer science, in 1984, from the California State University at Sacramento, Sacramento, CA, USA, and the M.S. degree in computer science with a concentration in symbolic systems from Stanford University, Stanford, CA, USA, in 1988. In 1998, he received the Ph.D. degree in intelligent systems from the University of Pittsburgh, Pittsburgh, PA, USA.

He worked as a Postdoctoral Fellow at the Robotics Institute at Carnegie Mellon University. He is currently a Professor of Computer Science and University Faculty Scholar at North Carolina State University, Raleigh, NC, USA, where he directs the Digital Games Research Center. His research interests span the computational modeling of narrative, the use of artificial intelligence techniques within computer games and virtual worlds.

Dr. Young is a Senior Member of the Association for the Advancement of Artificial Intelligence and is an Association for Computing Machinery Distinguished Scientist.