Advances in Cognitive Systems 4 (2016) 227-245 Submitted 6/2015; published 6/2016

Question Answering in the Context of Stories Generated by Computers

Rogelio E. Cardona-Rivera[1] RECARDON@NCSU.EDU
Thomas W. Price TWPRICE@NCSU.EDU
David R. Winer DRWINER@NCSU.EDU
R. Michael Young YOUNG@CSC.NCSU.EDU
Liquid Narrative Group, North Carolina State University, Raleigh, NC 27695 USA

Abstract

The QUEST model of question answering is a computational-cognitive model designed to describe a person's mental narrative comprehension process, which predicts behavioral responses to open-ended questions about the narrative. The model is based around a QUEST knowledge structure, a directed graph that captures the story's events and the causal and goal-oriented relationships between them. We present a family of computational methods to transform an automated planning narrative knowledge representation that has been successfully used for computational narrative generation, to a QUEST knowledge structure. We then report an experiment to test the hypothesis that these plan-driven knowledge structures have a meaningful relationship with a story consumer's comprehension of a narrative. The results help advance a research agenda that uses cognitive states as the target of computationally generated narratives.

1. Introduction

Narratologists (Boyd, 2009; Herman, 2013) and artificial intelligence (AI) researchers (Gervás, 2009; Winston, 2012; Mueller, 2013) suggest that *narrative intelligence* (i.e., generation and comprehension) depends on competencies that distinguish us from our primate relatives. Stories play a foundational role in our cognition and they are ubiquitous, serving as a target of interpretation and as a framework to understand the world around us (Schank, 1990; Herman, 2013). Driven by recognition that the study of narrative is a worthwhile endeavor and that computational modeling is well suited to understand this complex human phenomenon (Simon, 1996), the field of computational models of narrative has progressed primarily on the two fronts that make up the enterprise of narrative intelligence: narrative generation and narrative comprehension (Mueller, 2013). This work synthesizes these threads, enabling comprehension-driven computational models of narrative generation, by presenting methods that map an automated planning model of narrative generation (Riedl & Young, 2010) to a question-answering based model of narrative comprehension (Graesser & Franklin, 1990).

1. The authors contributed equally to the work presented here and should be considered co-first authors.

Our computational account of narrative generation leverages automated planning to model aspects of reasoning about stories and discourse about them (Young et al., 2013). This planning-based knowledge representation is well-suited for modeling key structural aspects of narrative, explicitly representing the causal and temporal structures of a plot (Young, 1999). However, key to narrative generation is modeling of the mind as it makes sense of stories. As people consume a narrative, their story comprehension faculties project a fictional world (Gerrig & Wenzel, 2015) that influences how they expect the future of the narrative to unfold, and transitively, their cognitive and affective responses. Concordantly, authors design stories to affect their audience in specific ways (Bordwell, 1989). Because the fundamental design criteria for a narrative artifact rest in the cognitive and affective responses they prompt in human consumers, computational models of narrative must go beyond story structure (Gervás, 2009). Minimally, we care about generating narratives that are comprehended in terms of causal relationships and goal-oriented structures. The work presented here is a step in service of this directive: to generate stories that elicit a specific effect (i.e., cognitive state) in the human consumer, we evaluate how well the data structures used to drive a narrative's generation can themselves model the comprehension process of of the generated narrative.

For this, we map narrative planning data structures to a computational model of narrative comprehension: the QUEST *cognitive model of question answering* (Graesser & Franklin, 1990). QUEST describes a story consumer's narrative comprehension process as measured through their ability to answer questions. This model posits that as a person consumes a narrative, she constructs a QUEST *knowledge structure* (QKS), a mental model that can be manipulated symbolically to return answers to specific queries. A series of studies (Graesser & Murachver, 1985; Graesser & Franklin, 1990; Graesser et al., 1991) designed to validate QUEST as a cognitive model of question answering used manually generated QKSs for each story. In this work, we present a mechanism to transform narrative planning data structures to QKSs. While we cannot compare the automatically generated QKSs to the hand-generated ones directly, we present an experiment that demonstrates our QKSs have explanatory and predictive power like the hand-generated structures used in the original QUEST studies.

There are several challenges in mapping one representation to another, mostly borne of the fact that the narrative planning and QUEST models were developed independently. First, some interpretation is required in order to establish semantic equivalence between the data structures of one model and those of the other. Second, while the planning data structures have a clear semantic interpretation, the QUEST model leaves some of their semantics underspecified. There are many candidate mappings that could be developed to go between the models and each mapping might serve different explanatory uses. In this work, we present three mappings – two developed previously (Christian & Young, 2004; Riedl, 2004) and one novel that builds upon prior work – that transform narrative planning data structures to their corresponding QKSs. These let us contrast their performance vis-á-vis predicting question-answering mechanisms in human story consumers.

The contributions of this paper are thus threefold: (a) we describe three techniques (two previously described by others) that map a planning-based narrative knowledge representation to a cognitive structure (i.e., QUEST knowledge structure), (b) we present the results of an experiment that demonstrate that the automatically generated QKSs can predict human responses to questions in a manner comparable to the hand-authored QKSs used in the original experiment by Graesser et

Table 1. The Crown Heist Story, an automatically generated narrative rendered as text.

This is a story about how Ethan was holding the Crown, Ethan was safe, Frank was safe and the vault holding the Crown was closed.

In the beginning, Ethan was at the Secret Entrance. Ethan could hack. Frank was at the Prison area. Frank was imprisoned. Frank knew about the Crown. The Guard was at the Vault area. The Secret Entrance was an exit. The vault holding the Crown could be accessed from a computer at the Switch Room. The Guard was holding the Prison Key. The Prison Key can unlock the Prison area. The Crown was at the Vault area. The vault holding the Crown was closed. Ethan wanted Frank not to be imprisoned. Frank wanted Ethan to know about the Crown. Ethan went from the Secret Entrance to the Vault area. The Guard got distracted by the computer. Ethan stole the Prison Key from the Guard. Ethan went from the Vault area to the Prison area. Ethan unlocked the Prison area with the Prison Key and freed Frank. Frank told Ethan about the Crown.

Ethan went from the Prison area to the Switch Room. Ethan hacked open the vault storing the Crown. The Guard went from the Vault area to the Switch Room. Ethan went from the Switch Room to the Vault area. Ethan stole the Crown from the Vault area. The alarm siren was set off. Ethan went from the Vault area to the Secret Entrance. Ethan escaped the building. Frank went from the Prison area to the Secret Entrance. Frank escaped the building. The Guard closed the vault storing the Crown. The End.

al. (1991), and (c) we demonstrate that different mappings can capture different aspects of human question answering. Our work here thus helps advance a research agenda that uses cognitive states as the target of computationally generated narratives.

While a discussion of QUEST's relationship to other cognitive models of narrative comprehension is beyond the scope of this paper,[2] we restate the elements of QUEST that are critical for discussion in Section 2. We present the mappings of planning structures to QKSs in Section 3. Because our work covers mappings using the same kind of knowledge representation, our discussion is in the style of a chronological successive refinement. We begin by presenting the narrative plan representation in Section 3.1. We continue in Section 3.2 by discussing both (a) the first technique for mapping a narrative plan to a QKS by Christian and Young (2004), and (b) the results from an empirical validation of their technique. The second technique, developed by Riedl (2004), depends upon an expanded knowledge representation, which is discussed in Section 3.3. We conclude the review in Section 3.4 by discussing (a) a second technique developed for mapping Riedl's narrative plan to a QKS and (b) the results Riedl obtained from an empirical validation of his technique. In Section 3.5 we present a novel approach that takes Riedl's expanded representation and computes a different QKS. In Section 4, we present an empirical evaluation for each of the three QKSs. Finally, in Section 5 we discuss how our work advances research on comprehension-driven computational models of narrative generation, as well as limitations of this work and directions for future work.

2. For comprehensive reviews, see Graesser et al. (1997) and McNamara and Magliano (2009).

Table 2. Types of nodes within QUEST knowledge structure graphs covered in this work.

Node Type	Description	Example
State	Describes a state assumed true until explicitly changed.	(Ethan has the jail key)
Goal	Describes a state or event desired by an agent.	(Ethan wants to free Frank)
Event	Describes an intended or unintended state change.	(Ethan pickpockets the Guard)

2. QUEST: A Cognitive Model of Question-Answering

As described by Graesser and Franklin (1990), QUEST is a cognitive model of question answering. More broadly, it is a computational-cognitive model (Sun, 2008), since it imputes a computational representation and procedure to the mental processes involved in the comprehension of narrative. Their system assumes that semantic content exists in the mind of a person as an *information source*, called a QUEST knowledge structure (QKS), which is manipulated symbolically to return answers to specific queries. The model also assumes, as we do, that a person's mental model of the situations in a story are propositional in nature (Johnson-Laird, 1983). Comprehension in QUEST is operationalized via an experimental question-answering paradigm: Their aim was to describe the narrative comprehension process by accounting for how human adults normatively answer certain classes of open-ended questions in story contexts. The question categories that QUEST can reason about are *Why?*, *How?*, *When?*, *What enabled X?*, and *What are the consequences of X?* questions. We also refer to the last two types as *enable* and *consequence* questions, respectively.

Table 3. Types of connecting arcs within QUEST knowledge structures covered in this work. Arcs go between types of nodes, where G denotes a *goal node*, and NG denotes a *non-goal node*.

Arc Type	Description	Example
Consequence	$\{NG\} \xrightarrow{C} \{NG\}$, where the source node causes or enables the sink node.	(Ethan pick-pockets the Guard) $\xrightarrow{C}$ (Ethan has jail key)
Reason	$\{G\} \xrightarrow{R} \{G\}$, where the sink node is a reason, motive, or super-ordinate node of the source node.	(Ethan wants to pick-pocket the Guard) $\xrightarrow{R}$ (Ethan wants to free Frank)
Outcome	$\{G\} \xrightarrow{O} \{NG\}$, where the sink node specifies whether or not the source node is achieved.	(Ethan wants to pick-pocket the Guard) $\xrightarrow{O}$ (Ethan pick-pockets the Guard)
Initiate	$\{NG\} \xrightarrow{I} \{G\}$, where the source node initiates or triggers the sink node.	(Ethan hacks open the vault) $\xrightarrow{I}$ (The Guard wants the vault to be closed) [not depicted in Figure 1]

QUEST is used to predict the goodness of each answer to a given question. Questions and answers are drawn from nodes in the QKS; a pair of nodes form a question-answer pair, with a designated question node and answer node. One can form these pairs can be formed from arbitrary pairs of nodes, but not all answers will serve as good – or even correct – answers to the questions. For

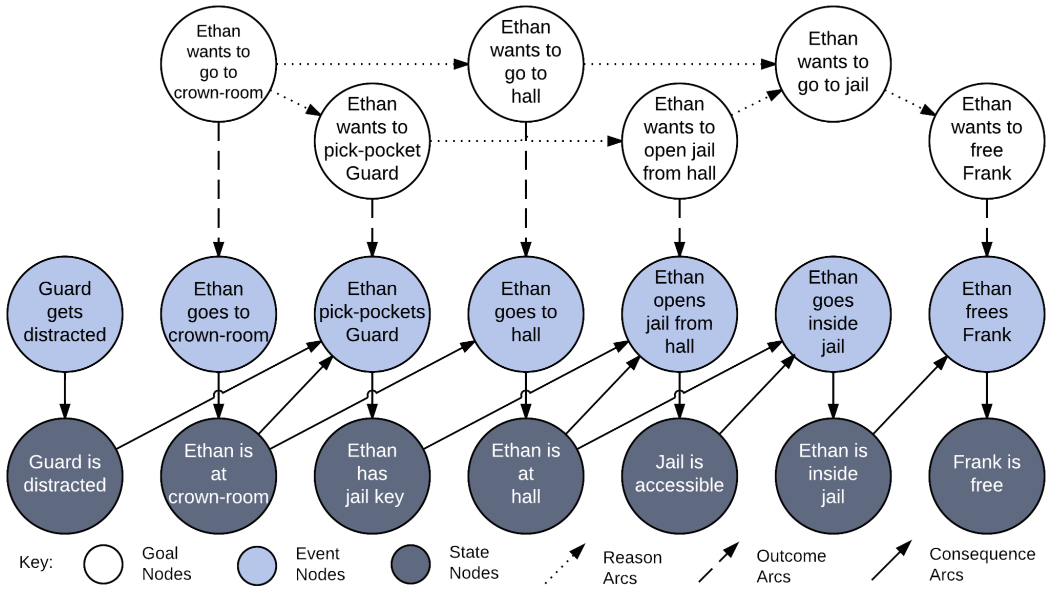

Figure 1. A QUEST Knowledge Structure for the first half of The Crown Heist Story.

each pair of nodes, QUEST predicts a *goodness-of-answer* (GOA) judgment, with answers labeled as *bad*, *possibly acceptable*, *moderately good*, or *very good*.

To facilitate the subsequent discussion of QKSs, we use a narrative called *The Crown Heist Story*. This story was constructed automatically and rendered as a film sequence (discussed in Section 4) for our experiment. Table 1 presents this story, realized as text,[3] to serve as a working example. A QKS is computationally represented as a directed graph of *statement* nodes. Figure 1 shows a QKS for the first half of The Crown Heist Story, from its beginning until the moment Ethan frees Frank, who is imprisoned at the start.

Statement nodes within a QKS contain either a simple sentence (e.g., as ⟨*subject*, *verb*, *simple-predicate*⟩), or a combination of such sentences (e.g., "x saw ϕ" where ϕ is of the form ⟨*subject*, *verb*, *simple-predicate*⟩). Both the statement nodes and their connecting arcs are typed based on their meaning and purpose. The three techniques we present to map planning structures to QKSs do not cover all types of QKS nodes and arcs. Tables 2 and 3 specify the nodes and arcs we cover, respectively, but Graesser and Franklin (1990) and Graesser et al. (1991) discuss the full range of structures.

2.1 Question Answering in QUEST

The QUEST answering procedure predicts the semantic and conceptual content of answers to questions given these inputs: an *information source* (i.e., a QUEST knowledge structure), a *question category*, and a *question focus*. Given a query, the procedure first determines the query's question

3. This text was realized via a straightforward automated mapping of data structures to sentences. Since we are not making claims about this realization, we will not discuss this mapping here.

category, and then its corresponding question focus. For example, for the QKS in Figure 1, we can ask the question: "Why did Ethan pickpocket the Guard?" In this instance, the information source is the QKS itself, the question category *Why*, and because this is a query over an intentional action, the question focus is the pair of nodes labeled "Ethan pickpockets the Guard" and "Ethan wants to pick-pocket the Guard." The procedure then searches the QKS according to the QUEST *convergence mechanisms*, narrowing the node space to a set of nodes that could serve as good answers to the question. QUEST incorporates three convergence mechanisms, but one (*constraint satisfaction*) dealt with answers that referenced story content consistent with the portrayed story world, but never explicitly shown in the narrative.[4] Since we do not account for such non-explicit story material we restrict our attention to the two remaining convergence mechanisms:

1) The *arc-search procedure*, which traverses the QKS starting from the question focus node and identifies candidate answer nodes to the query based on the question's type. There are unique arc-search procedures for each kind of question (Graesser & Franklin, 1990), but in all cases it first identifies which arcs are traversable and in what directions. For example, for *Why* questions, the procedure searches the QKS for both *superordinate nodes*, which are reachable from the question focus node through *Reason* arcs, and *goal initiators*, which connect to the question focus node or the superordinate nodes through *Initiate* arcs. In other words, according to the QUEST model, a good response for a *Why* question is either a goal or superordinate goal to which the question is tied, or an event that prompted the goal, or superordinate goal, in the first place. For the question "Why did the Ethan pickpocket the Guard?" one searches with these criteria from the question focus nodes "Ethan pickpockets the Guard," and "Ethan wants to pickpocket the Guard." Since no *Initiate* arcs appear in Figure 1, one can never find goal-initiator nodes. From the node "Ethan pickpockets the Guard," one can get to no other nodes through *Reason* arcs. However, from "Ethan wants to pickpocket the Guard," we can reach the nodes "Ethan wants to open jail from hall", "Ethan wants to go to jail", and "Ethan wants to free Frank" through *Reason* arcs. Thus, these are candidate answers for the question: "Why did the Ethan pickpocket the Guard?"
2) The *structural distance metric*, which measures the number of arcs between the question focus node and the node that serves as the answer to the question. For the three legal answers found by the arc-search procedure, the respective structural distances are 1, 2, and 3. For nodes not identified by the arc-search procedure (i.e., illegal answer nodes), the distance is the shortest path from the question focus to the node via any arc. In general, structural distance correlates negatively with the goodness of an answer to a question (Graesser et al., 1991).

The arc-search and structural distance metric are combined to identify good answers to questions as defined by the QUEST model; Graesser and colleagues demonstrated that QUEST's judgments correspond to those of human raters via the experiment described in the next section.

4. As a person consumes a story, they will make inferences over actions and states of the world that are not explicitly narrated, but are necessitated to make the narrative coherent, and are enabled by the events that are explicitly narrated (Myers et al., 1987). In the original QUEST studies, these inferences were collected as more data for manually expanding a hand-authored QUEST knowledge structure, which is what we aim to produce automatically. Graesser et al. (1991) augmented the QKS with this inferred narrative information with the understanding that the *constraint satisfaction* convergence mechanism would identify these inferred nodes during the process of computing an answer to a question.

2.2 QUEST Evaluation Procedure

A full review of all the experiments conducted to validate the QUEST model is beyond the scope of this paper. Instead, we focus on the evaluation with the context closest to ours: answering questions about stories (Graesser & Murachver, 1985; Graesser et al., 1991). Since we have replicated the QUEST evaluation experiment, we defer full discussion of the procedure to Section 4, where we present our experimental design. Here we present only a broad overview of the evaluation procedure and recap the main findings relevant to our work. This procedure consisted of two phases. In Phase I, Graesser and Murachver (1985) generated question-answer pairs for the two stories they used across the studies. For each statement node that explicitly appeared in the story text, they generated questions from every question category and then compiled an answer distribution for each question-answer pair. In Phase II, Graesser et al. (1991) generated unique question-answer pairs for each question, using a pool of answers that included that question's answer set (compiled in Phase I), and those from neighboring statement nodes in the QKS. They asked subjects to provide, for each question-answer pair, either a binary goodness-of-answer (GOA) judgment (it is a good/bad answer to the question), or a 4-point Likert (2009) GOA judgment (it is a bad/possibly acceptable/moderately good/very good answer to the question). The authors measured the individual impact of QUEST's three convergence mechanisms on GOA ratings for all question types. We will not discuss all the hypotheses and results of these studies, but the critical findings for our work include the relationship between the convergence mechanisms and GOA ratings for question types. Specifically, for all question categories (*Why*, *How*, *Enable*, *When*, and *Consequence*), if the question-answer pair passed the arc-search procedure, the model predicted a good answer. Similarly, if the question-answer pair failed the arc-search procedure, it predicted a bad answer. In addition, for *Why*, *How*, and *Enable* questions, structural distance was a significant predictor of GOA only when the question-answer pair failed the arc-search procedure.

3. Mapping Plan Structures to Quest Knowledge Structures

Prior work (Christian & Young, 2004; Riedl, 2004) has demonstrated a correspondence between partial-order causal link planning structures (Weld, 1994) and QUEST knowledge structures. That body of work sought to demonstrate that partial-order causal link (POCL) plans, which can adequately characterize key structural properties of narratives (Young, 1999), could additionally serve – through an appropriate mapping – as plausible models of a narrative consumer's understanding of a narrative. The mapping takes as input a POCL-like plan, and outputs the corresponding QKS structure meant to represent a human's mental structures after successful comprehension of a narrative. A mapping is evaluated based on how well the output QKS adequately predicts behavioral responses in the experiment used to validate the QUEST model itself (discussed in Section 4). In this paper, we summarize and expand on prior work in this area by introducing a mapping that refines prior efforts and by running an experiment to evaluate the mapping techniques. The three methods are similar in nature, and we provide them collectively in Table 4 to facilitate comparison. The three methods each use POCL plan structures, but both the mapping by Riedl (2004) and our mapping operate over an expanded POCL knowledge representation, called IPOCL, discussed in Section 3.3.

Because POCL is a subset of IPOCL, Table 4 shows them all as taking an IPOCL plan as input. We defer the presentation of the method to introduce the relevant data structures.

3.1 Partial-Order Causal Link Planning

A POCL planner finds a sequence of steps that transform some initial state of the world to some goal state of the world. A *world state* is conjunction of logical function-free ground literals that describe what is true and false in a story world. States are either established by the execution of a *step* or are true in the initial state. Steps are instantiated from templates called *operators*. Formally:

Definition 1 (Step) *A* step *is a tuple* $\langle T, P, E, L\rangle$ *where* T *is the type of the instantiated operator (e.g.,* PICK-UP*);* P *is a set of* preconditions*, literals which must be true before the step can execute;* E *is a set of* effects*, literals made true by the step's execution; and* L *is a label which distinguishes this step from other instances of the* T *operator.*

The set of all available operators (which when instantiated become steps) is called the *planning domain*. For generality, P and E can have variable terms to convey ideas such as "creature x steals item y". The assignment of a value to a literal with variables is recorded in a *binding*:

Definition 2 (Binding) *A* binding *is a tuple* (X, Y, D)*, where* X *is a variable term of a step in a partial plan;* Y *is either a variable term or a constant term in the quantification domain of* X*; and* D *is the* designation relation *between* X *and* Y*, and is either the* codesignation relation *(i.e.,* $X = Y$*) or the* non-codesignation relation *(*$X \neq Y$*).*

Plan steps are partially ordered with respect to time (Penberthy & Weld, 1992):

Definition 3 (Ordering) *An* ordering *is a tuple* $\langle s, u\rangle$ *where* s *and* u *are steps. The ordering specifies the relative order of the tuple's steps. We denote an ordering over two steps as* $s \prec u$*, where* s *must be executed before* u*.*

The POCL-planning process is one of least-commitment iterative refinement; in each iteration, the process records the step it has chosen to satisfy a condition in the plan and the reason for choosing that step. To record this dependency, it uses a data structure that encodes causal relations. Formally:

Definition 4 (Causal Link) *A* causal link *is denoted* $s \xrightarrow{p} u$*, where step* s *has an effect* p *and* p *is a precondition of step* u*. A causal link* $s \xrightarrow{p} u$ *implies the ordering* $s \prec u$*. Step* u*'s* causal parents *are all steps* s *such that there exists a causal link* $s \xrightarrow{p} u$*. A step's* causal ancestors *are the steps in the transitive closure of the parent relation.*

A POCL *planning problem* is defined by the initial state specification, goal state specification, and the planning domain. The solution to a POCL planning problem is a plan:

Definition 5 (Plan) *A* plan *is a tuple* $\langle S, B, \prec, L\rangle$ *where* S *is a set of steps;* B *a set of variable bindings;* $\prec$ *a set of orderings; and* L *a set of causal links. A* complete plan *is guaranteed to achieve the goal from the initial state. A plan is complete if and only if:*

- *For every precondition p of every step $u \in S$, there exists a causal link $s \xrightarrow{p} u \in L$ (i.e., every precondition of every step is satisfied).*
- *For every causal link $s \xrightarrow{p} u \in L$, there is no step $t \in S$ which has effect $\neg p$ such that $s \prec t \prec u$ is a valid ordering according to the constraints in $\prec$. In other words, it is not possible that a causal link can be made undone before it is needed.*

3.2 Christian and Young's (2004) Mapping

Christian and Young (2004) developed a technique to convert the POCL plan representation to a QKS representation. The mapping of POCL plans to QKSs is not straightforward due to the expressivity of the QKS and the relative precision of POCL plans. Nevertheless, one can map a standard POCL plan structure into a simple yet functional QKS; the generated QKS is limited to containing only Outcome, Consequence, and Reason arcs, and it only describes the actions of a single character. Their technique in Table 4 maps steps to Event and Goal node pairs and effects to State nodes. Event nodes are linked to their effect States by Consequence arcs, and Causal links are translated into Reason and Outcome arcs between steps' Event and Goal nodes.[5]

Christian and Young evaluated their mapping using a cinematic narrative in a game-based virtual environment (i.e., a film made in a video game) created to convey the plot of a narrative POCL plan. The plan was converted into a QKS to predict viewers' understanding of the cinematic. The authors followed an experimental procedure modified from that used by Graesser et al. (1991): 15 subjects were familiarized with the game interface and actions inside the game world, viewed the cinematic, and later rated *Why*, *How*, and *What enabled* question-answer pairs for their goodness of answer (GOA). The arc-search and structural distance predictor variables were calculated for each question-answer pair using the QKS generated from the POCL plan and constraint satisfaction was calculated by a novel method rather than the experts that Graesser et al. used. Christian and Young found arc-search and constraint satisfaction to be significant predictors of GOA rating for all question types and structural distance to be significant for *What enabled* questions.

3.3 Intentional Partial-Order Causal Link Planning

Riedl (2004) expanded the base POCL representation to generate story plans that explicitly reason about the apparent believability of characters. He demonstrated that audiences find characters in stories more believable when they execute steps directly in service of goals that they adopt during the development of a narrative arc. To that end, Riedl and Young (2010) defined a data structure on top of the base POCL representation that reified character intentions:

Definition 6 (Frame of Commitment) *A frame of commitment is a tuple $\langle S', P, a, g_a, s_f \rangle$ where S' is a subset of steps in some plan P, a is a character, g_a is some goal of character a, and s_f is a final step which has effect g_a. The steps in S' are all the steps that character a takes in order to achieve goal g_a. All steps in S' must be causal ancestors of s_f, and all steps in S' must be ordered before s_f.*

5. The original mapping contains an extra step for creating Goal nodes corresponding to the authorial goals in the planning problem. However, we use these nodes to represent *character* goals only, not authorial goals.

Table 4. The mapping method takes as input an IPOCL plan and output a QKS. Differences between mappings are shown on different sides of the vertical lines.

METHOD: IPOCL→QKS
INPUT: An IPOCL Plan (P := $\langle S, B, \prec, L, I \rangle$)
OUTPUT: A QUEST Knowledge Structure

1. Create a total ordering for the steps in P that is consistent with the ordering constraints $\prec$.
2. For each plan step $s_i \in S$:
   2.1 Create an Event node, ε_i.

(Ours)	*Christian and Young (2004)*; *Riedl (2004)*
2.2 $\forall$ Frames of Commitment $(c_j := \langle S', P, a, g_a, s_f \rangle) \in I$, where $s_i \in S'$:	2.2 Create a Goal node, γ_i.
2.2.1 Create a Goal node γ_{ij}.	2.2.1 Connect $\gamma_i \xrightarrow{O} \varepsilon_i$.
2.2.2 Connect $\gamma_{ij} \xrightarrow{O} \varepsilon_i$.	

   2.3 $\forall$ effects e of s_i:
      2.3.1 Create a State node σ_e.
      2.3.2 Connect $\varepsilon_i \xrightarrow{C} \sigma_e$.

3. For each causal link $s_j \xrightarrow{p} s_k \in L$:
   3.1 Create a State node σ_p for the effect p.
   3.2 Create an Event node ε_k and Goal node γ_k for the step s_k.
   3.3 Create a Goal node γ_j for the step s_j.

(Ours)	*Christian and Young (2004)*; *Riedl (2004)*
3.4 If $\exists$ Frame of Commitment $(c := \langle S', P, a, g_a, s_f \rangle) \in I$, where $s_j, s_k \in S'$, **or** if s_k is unmotivated, Connect $\sigma_p \xrightarrow{C} \varepsilon_k$.	3.4 Connect $\sigma_p \xrightarrow{C} \varepsilon_k$.

Ours; Riedl (2004)	*Christian and Young (2004)*
3.5 If $\exists$ Frame of Commitment $(c := \langle S', P, a, g_a, s_f \rangle) \in I$, where $s_j, s_k \in S'$, Connect $\gamma_j \xrightarrow{R} \gamma_k$.	3.5 Connect $\gamma_j \xrightarrow{R} \gamma_k$.

4. $\forall$ Frames of Commitment $(c := \langle S', P, a, g_a, s_f \rangle) \in I$, where c is motivated by step s_m, let ε_m be the Event node that represents s_m and γ_f be the Goal node that represents s_f in the QKS.
   Connect $\varepsilon_m \xrightarrow{I} \gamma_f$.

A character, or actor, resolves to accomplish those goals that they *intend*. The set of all characters is denoted by A.

Definition 7 (Intention) *An* intention *is a modal predicate of the form $intends(a, g_a)$ where a is an actor and g_a is a literal that actor a is committed to make true. A* motivating step *is a step which causes an actor to adopt a goal and that has an intention as one of its effects. A* final step *is a step that achieves some actor goal by having g_a as one of its effects.*

Steps that causally link a motivating and final step make up a frame of commitment. Given these new elements, we introduce a new data structure that builds on Definition 5:

Definition 8 (IPOCL Plan) *An IPOCL plan is a tuple $\langle S, B, \prec, L, I \rangle$ where S, B, $\prec$, L are as defined in Definition 5. I is a set of frames of commitment. An IPOCL plan is complete if and only if it is a complete POCL plan, and for every step $s \in S$, and every character $c \in A$, there exists a frame of commitment $f = \langle S', P, a, g_a, s_f \rangle$, such that $s \in S'$, and $c = a$.*

3.4 Riedl's (2004) Mapping

Riedl updated Christian and Young's mapping technique to work with IPOCL plan structures. The new mapping shown in Table 4 takes advantage of the fact that IPOCL plans incorporate a representation of character goals, motivations for these goals, and actions taken in their service. Reason arcs are therefore only drawn between Goal nodes that share a Frame of Commitment, and Initiates arcs are drawn between a Frame of Commitment's motivating step and its top-level goal.[6]

Riedl followed a similar protocol to Christian and Young (2004), but focused solely on *Why* questions, hypothesizing that these questions would better explained by a notion of intention. Additionally, he used generated text, rather than a film sequence, to communicate the narrative of his plan, and he used only the arc-search predictor variable. Riedl's primary hypothesis was that the use of IPOCL as an underlying plan structure would increase the QKS' predictive power, so he compared the results of two QKSs, one generated using a POCL plan structure of the story and another that used an IPOCL plan structure. The arc-search procedure for both QKSs accurately predicted GOA rating, reaffirming Christian and Young's hypothesis. The study also demonstrated that using IPOCL significantly improved the predictive power of the QKS for *Why* questions.

3.5 A New Mapping

As shown above, the plan representation and the mapping technique used when creating a QKS from a POCL plan can have an important effect on the resulting structure and its ability to predict GOA. To explore the ways that decisions in the mapping procedure can affect the resulting QKS, we created a third mapping based on Riedl's, but which has a stricter criterion for the use of Consequence arcs, as shown in Table 4. In a QKS, a Consequence arc between two statement nodes, A and B, indicates that "A causes or enables B" (Graesser et al., 1991) and, in previous mappings, each causal link in a POCL plan is mapped to a Consequence arc. However, in the QKSs constructed by Graesser et al., Consequence arcs pointing to *intentional actions* always originated from within an agent's local goal hierarchy. This comprises the nodes that make up an agent's plan, including superordinate and

6. Riedl's original mapping omits State nodes, but we have retained them for consistency. For simplicity, we have also omitted a step of his method that deals with delegation of goals from one character to another, which we do not address in this work. For the complete method, see Riedl (2004).

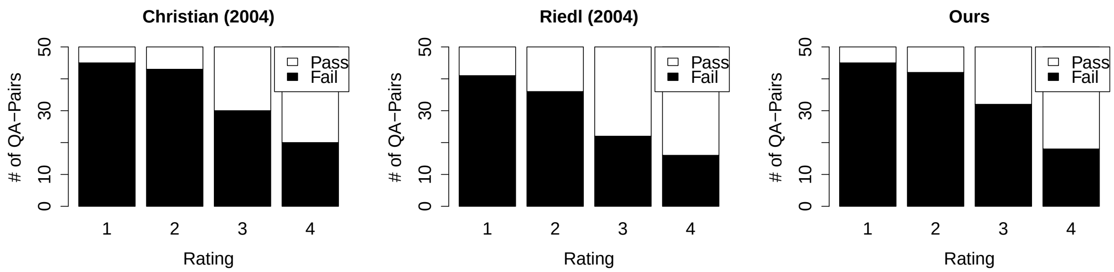

Figure 2. Stacked histograms for each mapping, showing for each GOA rating the number of QA pairs that passed the arc-search (in white) and failed the arc-search (in black) out of a random sample of 50 pairs.

subordinate goals and actions, reflecting the intuition that intentional actions are the consequences of an agent's plan. Causal links, however, often connect plan steps that are part of different frames of commitment and thus different goal hierarchies. Our mapping ignores these causal links, producing a QKS that resembles those created by Graesser et al. (1991). Additionally, our mapping does not create Goal nodes for unintentional actions, as these actions are not in service of a goal.

To understand the motivation for these changes, consider two steps from our example story: a) Ethan goes to the prison to free Frank and b) Ethan goes from the prison to the switch room to unlock the vault holding the Crown. In the POCL plan, there is a causal link between these two steps because going from the prison to the switch room requires Ethan to be at the prison. However, these two actions are part of two separate goal hierarchies (remember that Ethan did not learn of the Crown's existence until after freeing Frank). If we were to ask a question such as "Why did Ethan go to the switch room?" the set of appropriate answers should not include Ethan going to the prison to free Frank. But, if we connect these two actions with a Consequence arc, as in previous mappings (line 3.4 in Table 4), the arc-search will include Step 1, and transitively any of its causal antecedents, as appropriate answers, which our mapping does not. We hypothesize that this modified mapping will improve the resulting QKS's arc-search procedure for *Why* questions like this one.

4. Experiment

Graesser and colleagues (1985; 1990; 1991) generated the QUEST Knowledge Structures meant to encode a person's comprehension of a story manually using expert knowledge and evaluated how well they predicted behavioral results. We tested whether story plans generated by a computational process (i.e., an automated story planning system) could be automatically transformed into a QKS of comparable predictive power by replicating the two-phase experimental design outlined by Graesser et al. (1991). As described in Section 2.2, in the experiment we presented participants with question-answer node pairs (QA-pairs) from the QKS and asked them to give goodness-of-answer (GOA) ratings for for each pair. Similarly, for each QA-pair, we used the QKS to generate arc-search and structural distance predictor variables, as described in Section 2.1. We examined two hypotheses:

H1 A QKS generated automatically from a plan structure can be used to predict participants' comprehension of a narrative generated from that plan. Specifically:

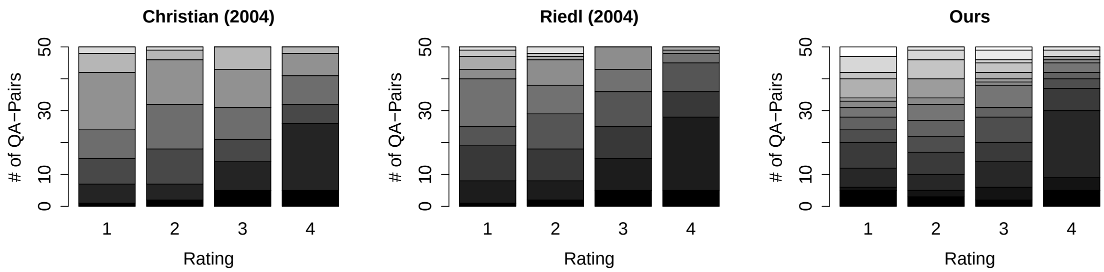

Figure 3. Stacked histograms for each mapping, showing for each GOA rating the distribution of QA pairs' structural distance, from closest (in black) to farthest (in white) for a random sample of 50 pairs.

a) The arc-search variable will have a linear relationship with participants' GOA ratings;
b) The structural distance variable will have an inverse linear relationship with participants' GOA ratings.

H2 Motivated changes to the mapping technique will produce improvements in the ability of the resulting QKS to predict GOA. Specifically:

a) Riedl's (2004) mapping will produce a higher quality model for *Why* questions than Christian and Young's (2004) mapping;
b) Our new mapping will produce a higher quality model for *Why* questions than Riedl's (2004) mapping.

To evaluate these hypotheses, we generated a narrative plan and a video[7] depicting the plan that we showed to subjects. We used the narrative plan to generate three different QKSs, one for each of the mapping techniques outlined in Table 4. We also developed an interactive demo to help each participant become acquainted with the logic of the world. All 40 participants completed the demo before beginning the experiment. After this they watched the narrative plan and answered questions about each event in the story. If possible, we tagged these answers to nodes in the QKS. As a result, each question-type (e.g., *Enabled*, *How*, *Why*) about each event in the story was linked to a set of possible answers. Using a procedure outlined in Graesser et al. (1991) in which some random nodes are also selected as possible answers, we created 912 QA-pairs consisting of a question and an answer. Later, a new set of 40 subjects watched the same film and judged the question-answer pairs. Each subject judged 200 QA pairs and furnished GOA ratings for a single QA pair via a 4-point Likert-type scale: (1) bad answer, (2) possibly an acceptable answer, (3) moderately good answer, or (4) very good answer.

4.1 Results

We collected 7552 GOA ratings for 780 unique QA pairs, averaging 9.7 ratings per pair. We converted these ratings to a numeric representation (1-4) but we treated them as ordinal data. Because we needed a single GOA rating for each QA pair, we took the mode of the ratings collected on a given QA pair as the representative rating for the pair. If a pair had ratings for which there

7. `https://youtu.be/Xy_4k4uuTrc`

Table 5. Results of the multiple ordered logistic regression on all data. S indicates significance at $p < 0.05$ and SC indicates significance at $p < 0.0033$, a corrected significance threshold for multiple (15) tests.

Mapping	Variable	*Why*	*How*	*What enabled*	*Consequence*	All
Christian and Young (2004)	Arc-Search (arc)	1.94_{sc}	2.00_{sc}	1.38_{sc}	1.52_{sc}	1.51_{sc}
	Distance (dis)	-0.268_{s}	-0.136	-0.194	-0.233_{s}	-0.222_{sc}
Riedl (2004)	Arc-Search (arc)	2.13_{sc}	1.67_{sc}	1.33_{sc}	1.61_{sc}	1.60_{sc}
	Distance (dis)	-0.419_{sc}	-0.137	-0.232_{s}	-0.302_{sc}	-0.268_{sc}
Ours	Arc-Search (arc)	2.29_{sc}	1.14_{s}	1.53_{sc}	1.82_{sc}	1.75_{sc}
	Distance (dis)	-0.086	-0.014	-0.001	-0.184_{sc}	-0.068_{s}

was no clear mode, we discarded it. We thus had 695 unique (QA pair, GOA rating) pairs. For each rating, we calculated the arc-search and structural distance variables for each of the QKSs generated by the three mappings discussed in Section 4.

Figures 2 and 3 show the relationship between the arc-search and structural distance predictor variables and the GOA ratings given by subjects. For clarity, the figures are based on a subset of the data, randomly sampled such that each (1-4) GOA rating was evenly represented with 50 samples, to compensate for an overrepresentation of low ratings in the data, although we performed all analyses on the entire data. Figure 2 highlights a trend that questions passing the arc-search are more likely to be rated highly by participants. Figure 3 shows a negative correlation between structural distance and GOA rating. The distribution of structural distance for low-rated GOA-pairs favors high values (lighter colors), while the distribution for high-rated pairs favors low values (darker colors).

4.2 Analysis

To evaluate hypothesis **H1**, we used a procedure based on previous work (Graesser et al., 1991; Christian & Young, 2004; Riedl, 2004). We carried out a multiple ordered logistic regression over the GOA ratings, with the arc-search (arc) and structural distance (dis) QUEST convergence mechanisms as independent variables:

$$GOA = \beta_0 + \beta_1(arc) + \beta_2(dis) + \epsilon \tag{1}$$

We stratified the 695 (QA pair, GOA rating) pairs on question type and used the same regression analysis for each question type. We used this stratification in order to isolate the effects of the QUEST convergence mechanisms on individual question categories. Table 5 outlines the results.

Analysis suggests that the arc-search variable is a significant predictor of GOA ratings for all mappings overall and nearly all question types. Structural distance is also generally significant across mappings, though less so both statistically and practically. It appears structural distance is only a significant predictor across question types for Riedl's (2004) mapping. This supports **H1**, as well as the hypotheses of Christian and Young (2004) and Riedl (2004), that a plan structure can produce a viable QKS which predicts the GOA ratings of questions-answer pairs pertaining to events of that plan. This serves to further support the choice of a POCL plan structure for narrative analysis and generation.

Table 6. AIC values for the multiple regressions in Table 5. For each column, a lower value indicates a better model for that question type. The last column indicates the overall model fit. Bold values are the lowest for a question type.

Mapping	*Why*	*How*	*What enabled*	*Consequence*	All
Christian and Young (2004)	359.55	**274.75**	321.56	349.91	1303.84
Riedl (2004)	**334.13**	280.89	**318.37**	338.18	**1269.61**
Ours	341.66	298.72	331.52	**330.09**	1302.43

To evaluate hypothesis **H2**, we compared the regression models produced by the three mappings discussed in Section 3. For each model, we computed the Akaike information criterion (AIC), a relative measure of model quality. Although AIC cannot determine the quality of the model in an absolute sense, we claim that all models adequately describe the data given the significant results in Table 5. Given a collection of models for a dataset, AIC estimates the quality of each model relative to each of the other models, and thus provides a metric for model selection (Burnham & Anderson, 2002). Table 6 shows the results.

Given the three candidate models, the best mapping, given by the minimum AIC, is Riedl's (2004). These results support **H2a**, along with Riedl's original hypothesis, that an IPOCL mapping generates a QKS that serves as a better predictor of *Why* questions' GOA ratings than one generated by Christian and Young's (2004) original mapping. Further, it seems this conclusion generalizes across most question types. To verify this, we compared the statistical model of Christian and Young's mapping to that of Riedl's mapping. We calculating the relative probability that the former model minimizes the information lost when using it to represent the underlying cognitive process that generated the collected data:

$$exp\left(\frac{AIC_{\text{(Riedl, 2004)}} - AIC_{\text{(Christian and Young, 2004)}}}{2}\right) \tag{2}$$

The above equation shows that, given our data and regression model, Christian and Young's mapping is $3.03 * 10^{-6}$ times as likely (relative to Riedl's) to minimize information loss for *Why* questions, and $3.69 * 10^{-8}$ times as likely to minimize information loss over all question types. Since the relative probabilities are so low, we can safely omit Christian and Young's model for future consideration on *Why* questions.

To evaluate **H2b**, we compared our mapping and that of Riedl (2004), which revealed our mapping as seemingly inferior. The reason is clear in Table 5, as the structural distance predictor variable fails to achieve significance for most question types. We hypothesize this is largely due to the fact that the QKS generated by our mapping is a disconnected graph, so structural distance values could not be computed for many of the QA pairs and were omitted, treated as a 0 in the regressions.[8]

8. We chose the value zero to represent the distance between two nodes that are unreachable from each other so as not to impact the regression. We also ran the analysis with infinity (or a very large number for use in regression), and this did not result in the structural distance variable reaching statistical significance.

Table 7. AIC values for models including only arc-search.

Mapping	*Why*	*How*	*What enabled*	*Consequence*	*All*
Riedl (2004)	353.35	**280.64**	**322.47**	346.12	**1300.05**
Ours	**342.97**	296.79	329.52	**339.38**	1307.47

Because we created our mapping the arc-search procedure in mind, we created a second set of regression models using only the arc-search predictor variable by dropping the $\beta_2(dis)$ term from Equation 1. The AIC values for these models appear in Table 7, although we have omitted the models themselves.

In this case, our mapping produces a model with minimum AIC for *Why* questions. Using Equation 2, we find that the model constructed from Riedl's (2004) mapping is $5.58 * 10^{-3}$ times as likely to minimize information loss on *Why* questions, partially supporting **H2b**; the arc-search procedure, not the structural distance metric, is improved for *Why* questions.

Although we have focused here on *Why* questions, our mapping differs on other question types as well. The arc-search variable for *Consequence* questions is improved, but the reverse holds for *How* and *Enabled* questions. While a post hoc analysis is not conclusive, a closer inspection of the mapping provides a plausible explanation. Recall from Section 3.5 that the new mapping assigns Consequence arcs to intentional actions only from within an agent's local goal hierarchy. This works well for *Why* and *Consequence* questions, which deal more with intentionality, but less well for *Enable* and *How* questions, which do not. In the example in Section 3.5, for instance, Ethan's going to the prison did *enable* him to later go to the switch room, even if it was not *why* he did it. Unfortunately, the arc-search procedures for *Why* and *Enable* questions are almost identical, and will yield many of the same nodes, even though these questions often evoke different answers from readers. Without modifying the QUEST procedures, it is more reasonable to expect a trade-off between question types than straightforward improvement.

5. Discussion and Conclusion

In this work we presented a family of methods that transform an automated planning data structure into an empirically supported computational-cognitive model of narrative comprehension (i.e., the QUEST cognitive model of question answering). Experimental analysis showed that all of the mapping variations predicted the goodness-of-answer ratings to questions about the story. Broadly, this supports the hypothesis that the underlying plan structures used to generate QUEST Knowledge Structures (QKSs) are related to a story consumer's comprehension, and that isomorphisms exist between the generative model and the descriptive model. Our work also presents a general two-step strategy for computational narratologists interested in evaluating narrative data structures that aim to achieve some cognitive effect. First, one defines a mapping from the narrative structures to a QKS and then one runs the experiment outlined in Section 4 to evaluate whether goodness-of-answer ratings are predicted from the narrative structures.

It is difficult to compare our results directly to those of prior experiments cited in this paper. We cannot, for instance, compare our procedurally generated QKSs to the hand-authored ones created by Graesser et al. (1991); we can only show that ours produce predictive models that achieve significance. This is due in part to the confounding role that the story itself plays in our analysis. Perhaps the strongest limitation of this work is the use of a single story, but our results are consistent with previous work (Christian & Young, 2004; Riedl, 2004) that used different ones. Our analysis cannot detect if the relative strengths and weaknesses of the different mapping methods would be consistent if we compared their predictions for other stories. The nature of specific events in the story may affect memory or inference processes in an idiosyncratic way. Therefore, future work that compares mapping techniques would benefit from demonstrating that differences between models are consistent across stories.

Two issues should also be considered when interpreting our results. First, there was an uneven distribution of GOA ratings in our data, with 455 of 695 rated as 1 (i.e., *bad*). This may be due in part to some participants' confusion about repeated story events and unclear location boundaries. Second, this work covers only a subset of the QUEST node and arc types, omitting Style nodes, Manner and Implies arcs, and several others.

The goal of this work has been to show that planning data structures are an effective proxy for a story consumer's comprehension and to lay the groundwork for more precise models of narrative comprehension and generation. We compared the mapping methods and found systematic differences between the resulting QKSs with respect to how well they fit subject data for particular question types. These results suggest that QUEST may benefit from further delineations in the node and arc types. As discussed in Section 4.2, the arc-search and structural distance procedures may make it difficult to achieve improvement for all question types. We believe that new predictor variables based directly from plan structures will produce more robust predictions of GOA ratings.

In general, we believe it is valuable to predict question-answering behavior directly from formalisms like those used in narrative planning. Novel formalisms will require new predictions and new arc-search procedures. For example, a recently developed plan-based model of conflict (Ware et al., 2014) suggests new question-answer relationships, such as ones that depend on hypothetical reasoning (e.g., *What if* ϕ *had occurred?*). Similarly, a model that makes claims about the availability of events in memory, e.g., that calculates event salience during comprehension (Cardona-Rivera et al., 2012), may predict answers to questions about what occurs next. Through works like these, we can extend QUEST's usefulness to the computational narrative community, as both an evaluation tool and for the purpose of intelligent control of narrative generation.

The ultimate aim of our research is to leverage computationally precise descriptions of cognitive states as the targets of narrative generation. By combining the narrative planning and QUEST models, we lend strength to narrative planning in terms of its representational power and generative capacity. Thus, an exciting area for future work is to use QKS during the generation process itself. By knowing how plan structures will translate into cognitive effects, a system can choose structures to achieve specific cognitive-related ends, enabling comprehension-driven computational models of narrative generation as a search through a cognitive state space.

Acknowledgements

The material is based upon work supported with funding from the NC State University Laboratory for Analytic Sciences, the U.S. Department of Energy (DE-FG02-97ER25308), and the National Science Foundation (CHS-1319912). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NC State University or those of any agency or entity of the United States Government.

References

Bordwell, D. (1989). *Making meaning: Inference and rhetoric in the interpretation of cinema*. Cambridge, MA: Harvard University Press.

Boyd, B. (2009). *On the origin of stories: Evolution, cognition, and fiction*. Cambridge, MA: Harvard University Press.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach (2nd ed.)*. Berlin: Springer.

Cardona-Rivera, R. E., Cassell, B. A., Ware, S. G., & Young, R. M. (2012). Indexter: A computational model of the event-indexing situation model for characterizing narratives. *Proceedings of the Third Workshop on Computational Models of Narrative* (pp. 34–43). Istanbul, Turkey: MIT Press.

Christian, D. B., & Young, R. M. (2004). Comparing cognitive and computational models of narrative structure. *Proceedings of the Nineteenth National Conference on Artificial Intelligence* (pp. 385–390). Menlo Park, CA: AAAI Press.

Fikes, R. E., & Nilsson, N. J. (1971). STRIPS: A new approach to the application of theorem proving to problem solving. *Artificial Intelligence*, *2*, 189–208.

Gerrig, R. J., & Wenzel, W. G. (2015). The role of inferences in narrative experiences. In E. J. O'Brien, A. E. Cook, & R. F. Lorch Jr. (Eds.), *Inferences during reading*, 362–385. Cambridge, UK: Cambridge University Press.

Gervás, P. (2009). Computational approaches to storytelling and creativity. *AI Magazine*, *30*, 49–62.

Graesser, A. C., & Franklin, S. P. (1990). Quest: A cognitive model of question answering. *Discourse Processes*, *13*, 279–303.

Graesser, A. C., Lang, K. L., & Roberts, R. M. (1991). Question answering in the context of stories. *Journal of Experimental Psychology: General*, *120*, 254–277.

Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology*, *48*, 163–189.

Graesser, A. C., & Murachver, T. (1985). Symbolic procedures of question answering. In A. C. Graesser & J. B. Black (Eds.), *The psychology of questions*. Hillsdale, NJ: Lawrence Erlbaum.

Herman, D. (2013). *Storytelling and the sciences of mind*. Cambridge, MA: MIT Press.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.

Likert, R. (2009). The method of constructing an attitude scale. In G. M. Maranell (Ed.), *Scaling: A sourcebook for behavioral scientists*. Piscataway, NJ: Transaction Publishers.

McNamara, D. S., & Magliano, J. (2009). Toward a comprehensive model of comprehension. *Psychology of Learning and Motivation*, *51*, 297–384.

Mueller, E. T. (2013). Computational models of narrative. *Sprache und Datenverarbeitung: International Journal of Language Processing*, *37*, 11–39.

Myers, J. L., Shinjo, M., & Duffy, S. A. (1987). Degree of causal relatedness and memory. *Journal of Memory and Language*, *26*, 453–465.

Penberthy, J. S., & Weld, D. S. (1992). UCPOP: A sound, complete, partial order planner for ADL. *Proceedings of the Third International Conference on Knowledge Representation and Reasoning* (pp. 103–114). San Francisco, CA: Morgan Kaufmann Publishers.

Riedl, M. O. (2004). *Narrative planning: Balancing plot and character*. Doctoral dissertation, North Carolina State University, Raleigh, NC.

Riedl, M. O., & Young, R. M. (2010). Narrative planning: Balancing plot and character. *Journal of Artificial Intelligence Research*, *39*, 164–167.

Schank, R. C. (1990). *Tell me a story: A new look at real and artificial memory*. New York: Charles Scribner.

Simon, H. A. (1996). *The sciences of the artificial (3rd ed.)*. Cambridge, MA: MIT Press.

Sun, R. (2008). Introduction to computational cognitive modeling. In R. Sun (Ed.), *Cambridge handbook of computational psychology*, 3–19. New York: Cambridge University Press.

Ware, S. G., Young, R. M., Harrison, B., & Roberts, D. L. (2014). A computational model of plan-based narrative conflict at the fabula level. *IEEE Transactions on Computational Intelligence and AI in Games*, *6*, 271–288.

Weld, D. S. (1994). An introduction to least commitment planning. *AI Magazine*, *15*, 27–61.

Winston, P. H. (2012). The right way. *Advances in Cognitive Systems*, *1*, 23–36.

Young, R. M. (1999). Notes on the use of plan structures in the creation of interactive plot. *Proceedings of the AAAI Fall Symposium on Narrative Intelligence* (pp. 164–167). Menlo Park, CA: AAAI Press.

Young, R. M., Ware, S., Cassell, B., & Robertson, J. (2013). Plans and planning in narrative generation: A review of plan-based approaches to the generation of story, discourse, and interactivity in narratives. *Sprache und Datenverarbeitung: International Journal of Language Processing*, *37*, 67–77.