Cognitive models of discourse comprehension for narrative generation

James Niehaus
Charles River Analytics Inc, USA

R. Michael Young
North Carolina State University, USA

Abstract

This article presents an approach to using cognitive models of narrative discourse comprehension to define an explicit computational model of a reader's comprehension process during reading, predicting aspects of narrative focus and inferencing with precision. This computational model is employed in a narrative discourse generation system to select and sequence content from a partial plan representing story world facts, objects, and events, creating discourses that satisfy comprehension criteria. Cognitive theories of narrative discourse comprehension define explicit models of a reader's mental state during reading. These cognitive models are created to test hypotheses and explain empirical results about reader comprehension, but do not often contain sufficient precision for implementation on a computer. Therefore, they have not previously been suitable for computational narrative generation. The results of three experiments are presented and discussed, exhibiting empirical support for the approach presented. This work makes a number of contributions that advance the state-of-the-art in narrative discourse generation: a formal model of narrative focus, a formal model of online inferencing in narrative, a method of selecting narrative discourse content to satisfy comprehension criteria, and both implementation and evaluation of these models.

Correspondence:
James Niehaus, Charles River Analytics, Inc.
625 Mount Auburn St.
Cambridge, MA 02138, USA
E-mail:
jniehaus@cra.com

1 Introduction

Narrative generation has long interested artificial intelligence (AI) researchers, at least partially because authoring and understanding narratives seems to be a function of higher intelligence. Narrative is often decomposed by researchers exploring the problems of generation into constituent pieces (Young et al., 2004; Szilas, 2007; Cheong and Young, 2008). A narrative may be divided between (1) the 'story' (or plot), the events of the narrative world and information about them; (2) the 'discourse', the ordering of those events as they are told to the audience; and (3) the 'medium' (e.g., film, novel, puppet show), the representation of those events as they are conveyed to the audience. One method for narrative generation is to first decide upon the elements in the story, then choose from those elements and order them to create a discourse, and finally realize the discourse within a specific medium. The second task, creating the discourse from the story, is of particular interest when considering narrative comprehension because the appearance and ordering of story elements in the

Literary and Linguistic Computing, Vol. 29, No. 4, 2014.

doi:10.1093/llc/fqu056 Advance Access published on 13 October 2014

discourse is critical to the process of comprehension.

Work by cognitive psychologists in the area of narrative discourse comprehension gives some insight into the mental models people use to understand narrative, including narrative focus and narrative inference. 'Narrative focus' is the salience of an element in the reader's mind. Narrative elements that were recently mentioned are generally read faster and more easily comprehended, as are elements that are highly related to the current situation (Kintsch, 1988, 1998; Zwaan et al., 1995; Langston et al., 1999). 'Narrative inference' is the process by which a reader adds information not presented in the narrative to his mental representation of the narrative. Strong evidence suggests that readers routinely construct inferences during reading. This process is often performed by the reader in the course of understanding the narrative with little or no conscious effort. Inferences may be classified as 'elaborative' or 'necessitated'. Elaborative inferences are inferences that increase the reader's information about the story world, but are not needed for comprehension. Evidence suggests that readers do not routinely make elaborative inferences. Necessitated inferences are those that aid in comprehending the narrative as told. Such inferences are constructed online (that is, during reading) by readers more frequently and reliably (McKoon and Ratcliff, 1992; Graesser et al., 1994).

Defining a computational model of online narrative comprehension is a significant problem that has yet to be solved. Cognitive models of narrative comprehension (e.g. McKoon and Ratcliff, 1992; Gerrig, 1993; Graesser et al., 1994; Kintsch, 1998; Zwaan and Radvansky, 1998; Langston et al., 1999) are not defined in sufficient rigor to be used to directly characterize computational comprehension algorithms. Narrative generation systems have developed formalisms for causality and intentionality and have employed reader models, but none have addressed the narrative focus and inferencing in online narrative comprehension, both of which are fundamental to narrative understanding.

This article makes a number of contributions that advance the state-of-the-art in narrative discourse generation, combined in a system named INferences For Extending Recall (INFER). These include:

(1) 'A formal model of narrative focus': This article presents a computational model for predicting online narrative focus in plan-based sequences of discourse content.
(2) 'A formal model of narrative inference': Also presented is a computational model for predicting online narrative inferences in plan-based sequences of discourse content.
(3) A 'method of selecting narrative discourse content to satisfy comprehension criteria'.
(4) 'Empirical evaluation of the above models': The above models are implemented in the INFER system and tested on the experimental data.

2 Related Work

There are four broad areas of research related to the work described here. First, because the goal of this work is to generate narratives to satisfy comprehension criteria, work in narrative generation is highly relevant. Narrative generation systems may roughly be divided between 'simulation systems', those that primarily simulate the narrative world, and 'deliberative systems', those that primarily deliberate over the choice of narrative elements and events to include in the narrative. INFER is a deliberative system; it chooses events to satisfy comprehension criteria. Previous deliberative generation systems (Lebowitz, 1984, 1985; Turner, 1992; Sgouros, 1999; Mateas and Stern, 2003; Szilas, 2003, 2007; Mott and Lester, 2006) differ mostly in representations of narrative structure and algorithms for the selection of narrative content; however, none focuses on sequencing of discourse content according to cognitive models of the audience.

Of these systems, several are similar in narrative representations to our approach. Riedl *et al.*'s Fabulist architecture (Riedl, 2004; Riedl and Young, 2004, 2005; Riedl and Sugandh, 2008) is a planning-based narrative generation system based on the intent-driven partial order causal link (IPOCL) narrative planner. Cheong's Suspenser (Cheong, 2007; Cheong and Young, 2008) selects discourse content

to form a narrative that maximizes feelings of suspense. Bae's Prevoyant (Bae and Young, 2008) uses foreshadowing and flashback to create surprise effects in narrative discourses. Although Suspensor and Prevoyant both select plan-based discourse content, none of these systems attempts to leverage online models of a reader's narrative focus and inferencing, and hence their aim and accomplishments are different than those of INFER.

Second, work in computational inferencing in narrative understanding is relevant. Many systems have been developed that perform aspects of automatic narrative understanding (DeJong, 1979; Norvig, 1987b; Hobbs et al., 1988; Shapiro and Rapaport, 1995; Frank et al., 2003), see (Ram and Moorman, 1999; Mueller, 2002) for reviews. Related approaches include Cullingford's Script-Applier Mechanism (SAM) (Schank, 1975; Cullingford, 1977) and Wilensky's Plan-Applying Mechanism (PAM) (1976) that apply hard-coded commonsense knowledge to the problem of inferring the relationships among events, characters, and intentions in narratives. Norvig's FAUSTUS (1987a) employs a general purpose marker passing mechanism to draw inferences about narratives. The Distributed Situation Space model by Frank et al. (2003) uses training data about sequences of situations to compute belief values for propositions in related situations. Mueller's model-based story understanding (Mueller, 2003, 2007) employs axioms in event calculus to model the changes in the situations of narratives. INFER's reader model is novel among these works because of its grounding in cognitive theories of narrative discourse comprehension.

Third, cognitive models of discourse comprehension form the theoretical basis of our approach. Cognitive psychologists have identified multiple layers of discourse structure used in narrative comprehension (Zwaan and Radvansky, 1998; Graesser et al., 2002). One of these levels, the situation model, is a mental representation of the narrative world state, including characters, settings, events in the plot, and the inter-relationships between these items (Graesser et al., 2002). INFER models the creation and use of the situation model. The situation model can be represented as a time-related series of individual models, one for each step of processing. Zwaan and Radvansky (1998) distinguish (1) the current model of the latest event, (2) the integrated model comprising all of the previous events, and (3) the complete model, which is the final model at the end of the narrative. When a new event is read, a current model is created to encapsulate the new information, and the integrated model is updated to incorporate the current model. The specifics of how the current and integrated models are represented and updated differentiate the theories of discourse comprehension.

In narrative focus, Kintsch (1998) developed one of the first general models; the Construction Integration (CI) model builds minimally structured associative nets out of the textbase propositions, extended in Langston et al.'s (1999) Connectionist Model (CM). Zwaan et al.'s Event Indexing (EI) model (1995) separates the possible dimensions of an event into space, time, causality, protagonists and objects, and intentionality. The assumptions of this model were validated in multiple experiments (Zwaan et al., 1995). Subsequent to the formalization of the EI model, an emerging body of research has confirmed, clarified, and extended the EI model (Rinck, 2000; Magliano et al., 2001, 2005; Mo et al., 2007; Therriault and Raney, 2007). INFER formalizes and implement aspects of the EI model for use in narrative generation.

In narrative inferencing, Graesser et al. (1994) present a Constructionist Theory of discourse processing, the primary principle of which is that the reader attempts a 'search (or effort) after meaning'. When readers comprehend a narrative, they strategically construct representations and reason with these representations to achieve a goal, the most common of which is sense making. This principle is broken into three critical assumptions.

- 'The reader goal assumption' states that 'the reader constructs a meaning representation that addresses the reader's goals'.
- 'The coherence assumption' states that 'The reader attempts to construct a meaning representation that is coherent at both the local and global levels'.
- 'The explanation assumption' states that 'The reader attempts to explain why actions, events, and states are mentioned in the text'. (Graesser et al., 1994)

The understanding of the production of inferences in narrative comprehension has been confirmed and expanded since the introduction of the Constructionist Theory (Linderholm, 2002; Allbritton, 2004; Trabasso and Wiley, 2005; Casteel, 2007; Shears et al., 2007). INFER embodies many of these theories and extends this work into a formal computational model of inference generation in narrative comprehension.

Fourth, this work employs partial-order planning (POP) (Penberthy and Weld, 1992) to construct plans for discourses, perform reasoning on the part of the reader, and generate new discourses. The IPOCL planner (Riedl and Young, 2004) forms the basis for our representation. Briefly, IPOCL's plan representation builds off of common POP models where a plan is composed of a set of primitive components. 'Steps' represent the actions in plan and are defined by an act type, a set of preconditions indicating the state of the world required for the step to execute correctly and a set of effects indicating the ways that the step's successful action changes the world. 'Variable bindings' link free variables in the steps to object constants in the domain. 'Ordering constraints' over the steps define a partial ordering on the steps' execution, and 'causal links' are directed arcs that run between pairs of steps in the plan to indicate that the source step changes the world in a specific way required for the successful execution of the destination step. IPOCL further defines 'frames of commitment', groupings of coherent steps in a plan that are all being performed in service of an intention to achieve one of the goals of the plan's characters. IPOCL searches the plan space by maintaining a list of plan 'flaws' for each plan. A flaw represents an issue that must be addressed for the plan to be considered complete, such as an unfulfilled precondition.

3 The INFER System: An Approach to Generating Discourse Plans that Satisfy Comprehension Criteria

3.1 Approach overview

The INFER system generates narrative discourses to satisfy comprehension criteria concerning narrative focus and inferencing. INFER incorporates theory from cognitive models and extends this theory to form a computational model of online narrative comprehension. INFER simulates the reading process by constructing a situation model representation, calculating narrative focus and predicting inferences in the form of causal chains of story world events. Relations between elements are represented in a semantic net. This formalization is inspired by the EI model (Zwaan et al. 1995). INFER models the process of online, bridging narrative inferences by generating sequences of events to either bridge gaps in causality or bridge gaps between the current world state and character intentions. This inferencing follows the central Constructivist Theory principle that readers constantly search for connection and meaning in the narratives that they read (Graesser et al., 1994). In generation, INFER selects discourse content from a partial plan expressing story world facts, objects, and events to achieve desired comprehension effects, either by shifting focus or prompting inferences.

3.2 Architecture

Figure 1 depicts the architecture of INFER. INFER constructs a 'Discourse Plan' to satisfy 'Comprehension Goals' from a story-level representation of the 'IPOCL Plan' and 'Planning Domain'. The 'INFER Algorithm' constructs the discourse plan through an incremental search of the space of possible discourses, satisfying criteria for 'Causal Inferencing', 'Intentional Inferencing', and 'Focus'. INFER uses the modified event indexing '(MEI) Reader Model' to make predictions for the reader's comprehension of the current discourse, and to choose elements that will prompt the desired comprehension effects. The MEI reader model incorporates a 'Reader's Plan' that simulates the reader's understanding of the causality and intentionality of the story and a 'Situation Model' semantic net that simulates the reader's focus. The resulting discourse plan is a totally ordered subset of objects, characters, intentions, and events from the story plan.

3.3 INFER input and output

The input to INFER is a source IPOCL plan, a IPOCL planning domain, and a set of

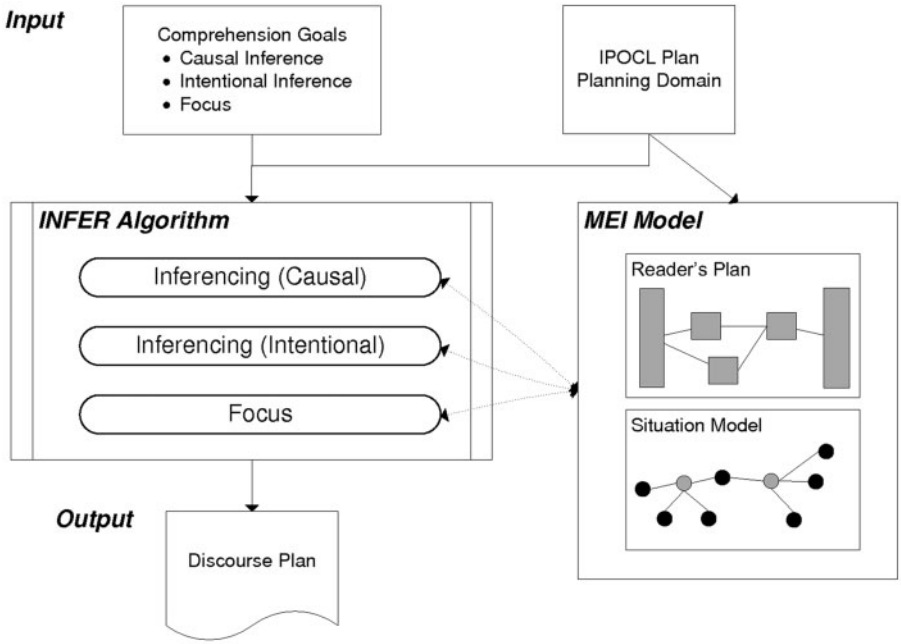

Fig. 1 INFER architecture

comprehension criteria divided into focus, causal inferencing, and intentional inferencing criteria. The output of INFER is a partially ordered discourse plan that may be linearized to form a sequence of discourse content, a total ordering over a subset of the elements in the source plan. The sequence of discourse content represents a proposition-level telling of a narrative which may be directly translated to a medium such as text or film.

Formally, a Discourse Plan is a set of steps, a set of binding constraints on the free variables in the steps, a set of ordering constraints on the steps, a set of causal links between steps, a set of frames of commitment, a subset of the steps that are causally prompted inferences, and a subset of the steps that are intentionally prompted inferences. A Sequence of Discourse Content is a total ordering over a subset of the plan steps, propositions, and frames of commitment of a Discourse Plan.

The sequence of discourse content, Figure 2 for example, can be used to create a narrative in a medium such as text or film by realizing each discourse element within the medium. Figure 3 shows

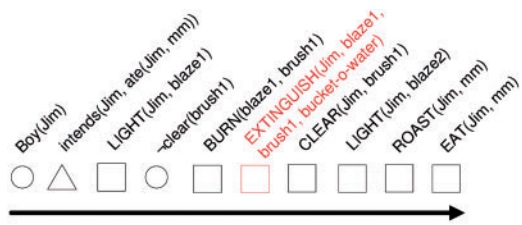

Fig. 2 A Sequence of Discourse Content. Events are squares, intentions are triangles, and propositions are circles

the translation from Figure 2 to text, performed by hand.

3.4 The MEI reader model

The two parts of the MEI model are (1) the 'Situation Model', a semantic network of nodes and edges to maintain the reader's shifting focus as new events are read, and (2) an IPOCL narrative plan to maintain the reader's understanding of the story world as the discourse is read. The situation

1. There was a boy named Jim.
2. Jim wanted to eat a marshmallow.
3. He lit a fire,
4. but the surrounding brush was not properly cleared!
5. The brush caught fire.
6. Jim quickly put out the fire.
7. Then he cleared the brush out of the way.
8. He relit the fire,
9. and roasted his marshmallow.
10. He ate his marshmallow.

Fig. 3 The sequence of discourse content from Figure 2 realized in text. Realization produced by hand

model and the IPOCL plan include the events, orderings, causal links, the characters, character intentions, objects, and locations of the narrative world.

3.4.1 *Situation model*

The MEI processing of the discourse content mimics the reader's comprehension process as described by the EI model (Zwaan et al., 1995; Zwaan and Radvansky, 1998) and related models of narrative discourse comprehension (McKoon and Ratcliff, 1992; Graesser et al., 1994, 2002). The MEI Situation Model provides a formal definition of 'foregrounding'—the process by which old information is re-activated—and 'updating'—the process by which new information is incorporated.

The Situation Model is built incrementally from the steps in a sequence of discourse content. The following principles from cognitive models of discourse comprehension guide the formal model of focus:

(1) The recency and relatedness of elements determines their salience (Kintsch, 1988; Zwaan et al., 1995; Langston et al., 1999).
(2) The strength of the connection between two events is a function of the number of shared dimensions of those events (Zwaan et al., 1995).
(3) After reading each element, focus spreads from the new element to related elements in an amount proportional to the strength of connection between them (Kintsch, 1988; Langston et al., 1999).

The Situation model is defined formally as a weighted bipartite graph between event nodes and dimension nodes (whose type is one of space, time, causal, object, or intention). Updating the Situation Model proceeds in stages. First, a new model is created according to the type of the last element read:

- If the element is a proposition, a dummy event node is created and object nodes for the proposition and the arguments are created and attached to the event node.
- Similarly, if the element is an intention, a dummy event node is created and intention nodes for the intention and object nodes for each argument are created and attached to the event node.
- If the element is a plan step, the new model is created as follows:

(1) Create a new event node.
(2) Create a new space node, time node, and causal node, for the event.
(3) For each proposition in the preconditions and effects of the step, create a new object node. For each term in the proposition create a new object node.
(4) Create an intention node for each of the intentions to which the step is linked.
(5) The new links are all given the weight of 1.0.

For example, Figure 4 shows the transition from the plan step representing 'Jim puts out the blaze with a bucket of water' to the new situation model for this event.

Second, once the new model has been created, the integrated model is then updated by incorporating the new model and simulating shifts in focus. The initial integrated model is empty of nodes, and it is expanded over the course of reading by incorporating each new model as it is created. Figure 5 shows an example integrated situation model.

To incorporate the new model and simulate focus, all of the nodes and edges are combined

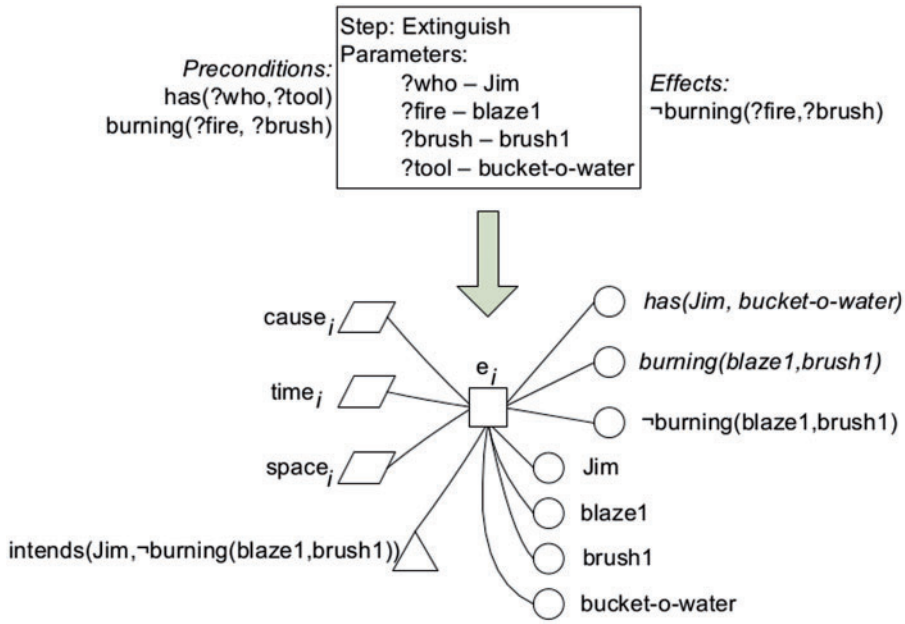

Fig. 4 A transition from a plan step to a new model. Event nodes are squares, intention nodes are triangles, proposition nodes are circles, and time, space, and causality nodes are rhombuses

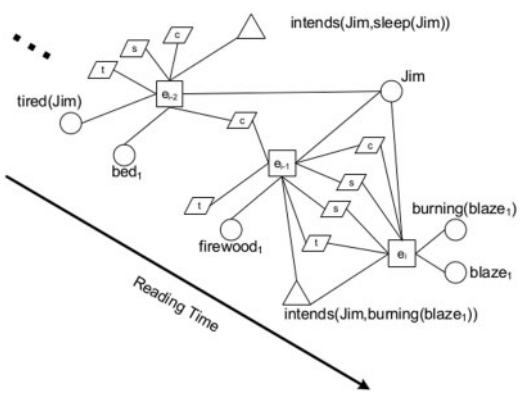

Fig. 5 A Situation Model during processing. Event nodes are squares, intention nodes are triangles, proposition nodes are circles, and time, space, and causality nodes are rhombuses

into a single graph. Object nodes are considered to be the same node if their propositions are the same or are codesignated; all other nodes are considered unique. Update the edge weights of the new graph by (1) discounting all existing weights by a constant percentage and (2) increasing the weights of the edges closest to the new event node. To increase the weights, the shortest-path-distance function determines the distance between the each edge and the event node. A simple step function, F, discounts the effects, as the edges are further away. Let v be a node adjacent to e along a shortest path to e_i with minimal degree, $D(v)$ be the degree of v, and c_s be a constant, then the spreading weight function, S, is defined as:

$$S(x, e, e_i) = \frac{F(SPD(e, e_i))}{D(v)} * c_s + x \qquad \text{(A)}$$

By which the weights on the edges closest to the new event node are increased the most, and then less as the edges become further. The increase is divided equally among vertices connected to the same node. Finally, the focus of any node in the network can be computed by approximating the value of this node in the Markovian steady state distribution.

To compute the ease of reading a new fact or event, the focus of the nodes to which it will be linked is averaged. This results in a 'relatedness score'. This method is used to predict reading times in the experiments described in Section 4.

3.4.2 *Inferencing*

Inferences predicted by the MEI model are causally related sets of events. They are represented as sets of IPOCL steps with causal links, ordering constraints, and binding constraints. To create inferences about the narrative world from the perspective of the reader, the events, characters, and intentions are represented as a partial IPOCL plan containing all of the steps, propositions, and frames of commitment in the input sequence of discourse content up to the current point of reading. To construct the IPOCL plan, each element in the input sequence of discourse content is processed in order:

- If the element is a step, add the step and all of its bindings, causal links, and preconditions to the plan.
- If it is a character intention, then add a new frame of commitment.
- If it is a proposition, create a dummy step with the sole precondition and sole effect of the proposition.
- Add an ordering from each step and frame of commitment preceding the element to the element.

For each new element processed, a copy of the reader's story is made, and causal and intentional inferencing, defined below, are employed to identify the new sets of steps that constitute the inferences. The MEI model's inferencing algorithms follow a small set of guiding principles that have been observed in psychological studies of discourse processing.

(1) Inferences in online comprehension must be necessitated and enabled (Graesser and Clark, 1985). That is, they must be required in the reader's search after meaning (as per Constructionist Theory (Graesser et al., 1994)) and they must not be overly difficult to construct.
(2) Readers more often make inferences about elements that are in focus (Graesser et al., 2002).
(3) Readers may make inferences that are incorrect. Predictions about unknown properties of a narrative may prove false, even when the possibilities are highly constrained.
(4) Inferences aid in comprehension (Graesser et al., 1994). Comprehension often requires the reader to fill in the gaps presented in the narrative, and inferences allow the reader to do so.

3.4.3 *Causal inferencing*

Causal inferencing occurs when the reader reasons about how events may occur or may have occurred to produce the current narrative situation. To satisfy the criteria of necessary and enabled, INFER only seeks to construct causal inferences when (1) there is a break in a causal chain in the Reader's Story and (2) the search for a complete sequence of events is short and there are few alternative sets of events. Once the inference is constructed, the salience of the elements related to the inference can be checked to satisfy enablement. The search algorithms are presented in Figures 6 and 7. The salience of an inference is the sum of the activations of its elements.

3.4.4 *Intentional inferencing*

As with causal inferences, intentional inferences are constructed when they are necessitated and enabled by the current context. Necessitation is achieved in this case by the statement of intent. The MEI model searches for intentional inferences beginning when an intention is presented in the discourse, and ending when the intention is realized. The search algorithms are presented in Figures 8 and 9. The salience of an inference is the sum of the activations of its elements.

3.5 *Generation*

This section presents a partial-order planning algorithm to generate a Discourse Plan by selecting and ordering content from a source plan. Generating a Discourse Plan is accomplished by a plan-space search in the space of partial plans. The search

Let A=L=$\emptyset$ and S_r be a sequence of plan steps.
Algorithm: CinfNec(A, L, S_r)

1. **Termination** If S_r is empty, return L.
2. **Add Propositions** Let s be the first step of S_r with effects E. Let A'=A, L'=L, $S'_r = S_r - s$. For each precondition p of s,
   - If $\neg p \in A$ then $L' = L' \cup \{p\}$
   - If $\neg p \notin E$ then $A' = A' \cup \{p\}$

   For each effect p_e in E, $A' = A' \cup \{p_e\} - \{\neg p_e\}$
3. **Recursive Invocation** call CinfNec(A', L', S'_r)

Fig. 6 Finding necessitated Causal Inferences

Let L be a set of propositions that have changed truth value in the Reader's Story P_r, and $I = \emptyset$

Algorithm: CinfEn(L, P_r, I)

1. **Termination** If L is empty, return the elements of I, P_i that pass the salience check (P_i, P_r) .
2. **Planning** Let p be the first proposition of L. $L' = L - \{p\}$. Let P_I be the set of complete plans obtained by conducting planning up to depth d with the initial state as P_r and the open precondition p as the only initial flaw. $I' = I \cup P_I$.
3. **Recursive Invocation** call CinfEn(L', P_r, I')

Fig. 7 Finding enabled Causal Inferences

Let $A = L = \emptyset$ and S_r be a sequence of plan steps.
Algorithm: IinfNec(A, L, S_r)

1. **Termination** If S_r is empty, return L.
2. **Add Propositions** Let s be the first element of S_r, $L' = L, A' = A, S'_r = S_r - s$.
   - If s is an intention, $L' = L' \cup \{s\}$.
   - If s is a step then for each precondition p, if $\neg p \notin E$ then $A' = A' \cup \{p\}$, and for each effect p_e, $A' = A' \cup \{p_e\} - \{\neg p_e\}$.
3. **Clean Intentions** $L' = L' -$ intentions which represent propositions in A.
4. **Recursive Invocation** call IinfNec(A', L', S'_r)

Fig. 8 Finding necessitated Intentional Inferences

begins with an initial root plan nodes and iteratively instantiates plan refinements as children nodes, which themselves may be recursively refined. Each plan in the space maintains a list of its flaws, and children plans in the search space are generated by choosing a flaw of a parent plan and applying the appropriate resolution method as specified in the algorithm defined below.

Let L be a set of unfulfilled intentions in the Reader's Story, P_r and $I = \emptyset$
Algorithm: IinfEn(L, P_r, I)

1. **Termination** If L is empty, return the elements of I, P_i that pass the salience check, $SalientInf(P_i, P_r)$.
2. **Planning** Let t be the first intention of L, p be the proposition of t and c be the character of t. $L' = L - \{p\}$. Let P_I be the set of complete plans obtained by conducting planning up to depth d with the initial state as P_r, the open precondition p as the only initial flaw, using only actions bound to the character c. $I' = I \cup P_I$.
3. **Recursive Invocation** call IinfEn(L', P_r, I')

Fig. 9 Finding enabled Intentional Inferences

The user specifies the initial causal inferencing criteria (i.e., what steps are to be causally inferred), intentional inferencing criteria (i.e., what steps are to be intentionally inferred), and focus criteria (i.e., minimum focus on specific elements), and each criterion is converted to a flaw in the root Discourse Plan. This algorithm generates plans such that the MEI model, described in Section 3.4, will compute the desired focus and inferences as specified in input comprehension criteria. The Discourse Plan is structured such that any valid linearization of this plan into a sequence of discourse content will satisfy the comprehension criteria. Therefore, generation rests on the predictions of the MEI model about reader comprehension.

Discourse Plans are generated in partial order, then linearized to form an output Sequence of Discourse Content. A Discourse Plan is an IPOCL plan (Riedl and Young, 2004) that contains steps, bindings, causal links, orderings, and frames of commitment. A Discourse Plan may also include inferred steps, which are not included in the final Sequence of Discourse Content. These inferred steps are used as place holders to ensure that inferences are enabled and necessitated. The final Discourse Plan must be constructed such that any linearization is a satisfying Sequence of Discourse Content.

Figure 10 is INFER's basic algorithm for generation; the formal algorithm specification can be found in Niehaus (2009). INFER selects a flaw and calls the appropriate routine to refine the plan to remove the flaw, possibly creating new flaws for the refined plan. Causal Inference flaws denote steps that are to be causally inferred. Intentional Inference flaws denote steps that are to be intentionally inferred. Link to intention flaws denote steps that are to be causally linked to an intention (for intentional inferencing necessitation). Open precondition flaws denote preconditions of inferred steps that are to be established in the plan (for inferencing enablement). Open effect flaws denote effects of inferred steps that are to be causally linked to noninferred steps (for causal necessitation). Lastly, focus flaws denote elements which are to have a minimum focus.

'Prompting Causal Inferences'. Discourse Plans with Causal Inference flaws are refined to include the inferred step, link the preconditions and one effect of the inferred step to a noninferred step, either directly or along a causal chain. Plans with Open Precondition flaws are refined to add a causal link to the precondition, similar to existing algorithms (Penberthy and Weld, 1992). Steps may be reused from the discourse plan, included from the source plan, or created as inferred steps. Chains of inferred steps must eventually have their preconditions satisfied and an effect satisfied by noninferred steps, linking the inference to what is told.

'Prompting Intentional Inferences'. Similarly, Discourse Plans with Intentional Inference flaws are also refined to include the inferred step, with the only difference from causal inferences being that an intentional frame is at the end of the causal chain, rather than another step. Discourse Plans with Focus flaws are refined by adding related steps to increase the focus of a particular element.

Algorithm: Refine Discourse Plan

Parameters: a Discourse Plan, a source plan, a list of flaws, and an action library

1. **Termination Check:** If there are no more flaws, return a non-deterministically chosen linearization of the plan.
2. **Flaw Selection and Refinement:** Non-deterministically select a flaw, either (a) a causal inference flaw, (b) an open precondition flaw, (c) an open effect flaw, (d) an intentional inference flaw, (e) a link intention flaw, or (f) a focus flaw. Switch on the flaw type:
   (a) **Prompt Causal Inference:** Add the step to be inferred. Create open precondition flaws for each of its preconditions and a single open effect flaw to ensure the inference is causally necessitated.
   (b) **Establish Open Precondition:** Either (1) create a causal link from an existing step in the plan to the flaw's precondition; (2) add a new step from the source plan and create a causal link from this new step to the flaw's precondition, include all of the ordering constraints for the new step from the source plan; or (3) add a new step to be inferred and create a causal link from this new step to the flaw's precondition, create open precondition flaws for each of the new step's preconditions.
   (c) **Link Open Effect:** Either (1) create a causal link from the flaw's step to another step in the plan; (2) add a new step from the source plan and create a causal link from the flaw's step to this new step, include all of the ordering constraints for the new step from the source plan; or (3) add a new step to be inferred and create a causal link from the flaw's step to this new step, create open precondition flaws for each of the new step's preconditions, and create an open effect flaw for this new step.
   (d) **Prompt Intentional Inference:** Add the step to be inferred as a place holder for further refinement. Create open precondition flaws for each precondition and a link to intention flaw to ensure the inference is intentionally necessitated.
   (e) **Link to Intention:** Either (1) create an intentional link from the flaw's step to an intentional frame or (2) add a new step, create a causal link from the flaw's step to the new step, create open precondition flaws for the new step, and create a new link to intention flaw for the new step.
   (f) **Shift Focus:** Add a new step from the source plan that contains the flaw's focus element; include all of the ordering constraints for the new step from the source plan.
3. **Threat Resolution:** Find any step that might threaten to undo any causal link or is redundant with an exclusive causal link. For every step, non-deterministically do one of the following:
   - **Promotion or Demotion**: If possible, add ordering links to move the threatened steps to occur before or after the threat in the plan.
   - **Separation**: If possible, add binding constraints on the steps involved so that no conflict can arise.
4. **Consistency Check:** If a refinement does not exist or is inconsistent, backtrack.
5. **Recursive Invocation:** Call the planner recursively with the refined plan and list of flaws

Fig. 10 Algorithm for refining a Discourse Plan to remove all flaws

'Threat Resolution'. The Discourse Plan is partially ordered, and the effects of causal links may be threatened to be undone by steps that may intervene. Discourse Plans have two types of causal links, exclusive and nonexclusive. An exclusive link keep other steps from destroying necessitation of an inference by requiring that the head step be the sole cause of the effect. Links created by open effect flaw refinements and link to intention flaw refinements are exclusive, and all other links are nonexclusive. Threats are handled by INFER in the standard method of promotion, demotion, separation, or backtracking (Penberthy and Weld, 1992).

'Pruning'. The prompting of inferences by the Causal Inference and Intentional Inference flaw refinements ensures the necessitation and logical enablement of the inference. However, an inference chain which is too long, or too difficult to construct does not meet enablement criteria. Because the causal chains only grow as the search proceeds, once the inference can be deemed too difficult to construct, that Discourse Plan can be pruned from the search space.

'Linearization'. Finally, if a plan is found that has no remaining flaws, the plan is linearized (any linearization will satisfy the comprehension criteria) via methods such as topological sort (McAllester and Rosenblitt, 1991) and returned as a solution.

3.6 Generation example

This section describes two examples of the generation algorithm operating on a source plan describing events in an old western town. First, Figure 11 shows the result of generating a discourse plan with a causal inference that the townsperson Sally was pickpocketed by the robber named Robbie. To generate this sequence, the refine discourse plan algorithm is instantiated with a source plan and a single causal inference flaw specifying the PICKPOCKET-1 step. To satisfy the open preconditions of PICKPOCKET-1, the algorithm includes the HATCH-PLAN-2, WITHDRAW-MONEY-4, MOVE-TO-ALLEY-3, and WALK-TO-5 steps and associated links from the source plan, creating exclusive causal links to the PICKPOCKET-1 step. To satisfy the open effect flaw, the FAIL-BUY-DRESS-6 step is included from the source plan. When this Discourse Plan is linearized and converted to a medium such as text, the PICKPOCKET-1 step is left out, allowing the reader to make the inference that Sally was pickpocketed to complete the causal chain between the rest of these steps.

Second, Figure 12 shows the result of generating a discourse plan with an intentional inference that Robbie will hold up the bank. Similar to the causal inference example, a placeholder step, HOLD-UP-BANK-7, is inserted and its preconditions are addressed by including steps and links from the source plan. However, to establish the intentional necessity, FRAME-1 is included stating the intention of Robbie to have the money, and a new inferred step, COLLECT-MONEY-FROM-HEIST-10, must be created to link the HOLD-UP-BANK-7 step to this intention. Additional links and steps are included to satisfy the preconditions of this new inferred step as well. Given descriptions of the steps in this discourse plan, a reader may make the two-step inference that Robbie will hold up the back and collect the money from the heist to satisfy his intention.

4 Experimental Evaluation

To evaluate INFER, we employed three separate experiments. The first experiment tests aspects of narrative focus, observing reading under conditions of varying levels of narrative focus as predicted by the MEI model. The second experiment tests the inference prediction of the MEI model, observing word recognition time under prompted and non-prompted inference conditions. The third experiment tests the ability of INFER to generate Sequences of Discourse Content to prompt inferences. The results of these experiments provide support for the computational model of narrative comprehension and generation presented in this work.

4.1 Experiment 1

This experiment tests aspects of the MEI model of narrative focus. Sentence and word reading times have been shown to relate to measures of focus

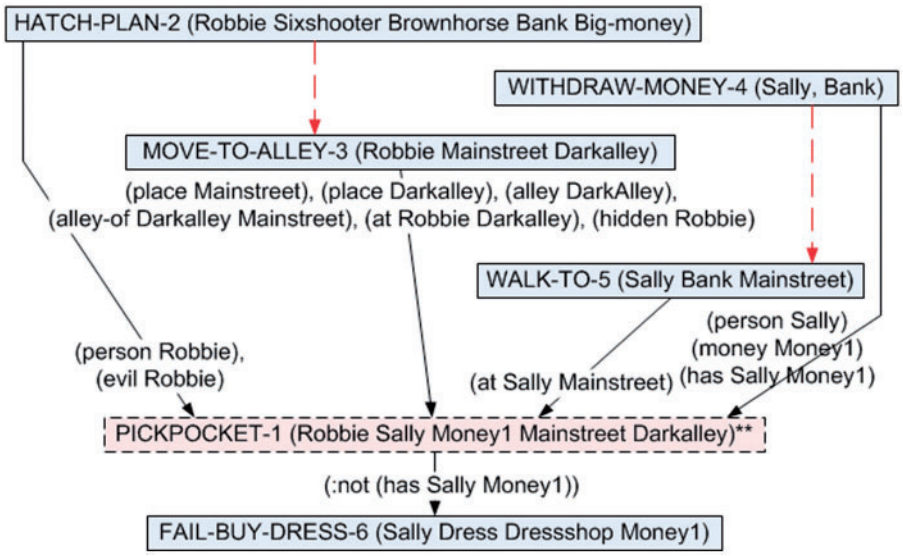

Fig. 11 Discourse plan with a necessitated and enabled causal inference of the PICKPOCKET-1 step

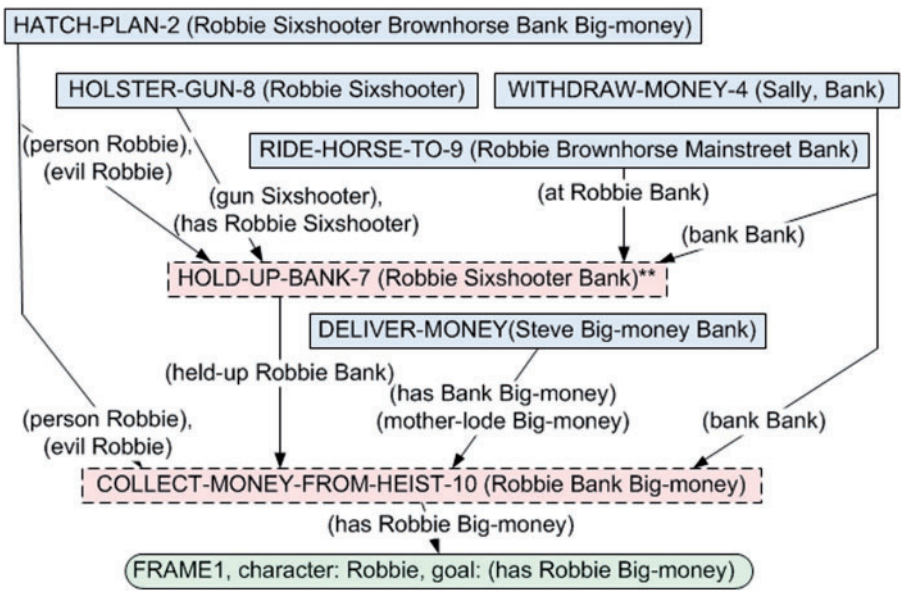

Fig. 12 Discourse plan with a necessitated and enabled intentional inference of the HOLD-UP-BANK-7 step

and activation (Kintsch, 1998; Graesser et al., 2002). The hypothesis for experiment 1 is that sentences that have higher relatedness will have shorter average reading times. To test this hypothesis, narratives were manually created to include sentences with various levels of relatedness and sentence reading times were recorded and regressed against the relatedness rating.

4.1.1 *Method*

Materials. The narratives for the first experiment consisted of four experimental narratives and eight filler narratives. The four experimental narratives were created by creating a complete IPOCL story plan by hand (generation was not tested by this experiment), manually selecting a Sequence of Discourse Content from the story plan, and generating text for the sequence using simple text templates. Each of the four experimental narratives had two conditions—high relatedness and low relatedness. In the high relatedness condition, a single test sentence was placed in the narrative such that it had a high relatedness score of greater than 5. In the low relatedness condition, this test sentence was swapped with a sentence that had a low relatedness score in that context of less than 3. The exact narratives can be found in the Appendix of Niehaus (2009). As discovered in piloting and previous protocols, the filler narratives reduced variability in the data by allowing participants more time to reach a steady flow of reading.

Procedure. The narratives were displayed one line at a time by a Java applet within a browser window. The experiment began with instructions to read for comprehension and a short example narrative. The participants were not informed they would be timed. The participants advanced the narrative one sentence at a time by pressing the left mouse button. Reading times were recorded after each sentence. At the end of each narrative two word recognition tests were administered to prompt readers to read the sentences without blindly clicking through; word recognition was not used in analyzing the data. The participant was presented with the question as to whether a particular word appeared in the narrative and indicated either 'yes' or 'no' by selecting an option with the mouse. There was a 500-ms pause between questions and between a question and the start screen for the next narrative. Participants were given untimed rest periods after the first two sets of four narratives.

Design and Subjects. Twenty-two middle school teachers participating in a technology research program volunteered to participate. Each participant only read one condition of each narrative. The narrative conditions were counter-balanced in a rotating sequence, and the order of narratives given to each participant was randomized.

4.1.2 *Results*

First, a bivariate regression analysis carried out to model how reading times vary with overall relatedness ratings (calculated for every sentence but the first of each narrative) found that relatedness was a small, but statistically significant, predictor of reading times across all sentences ($\beta = -0.175$, $\alpha = 2.98$), confirming the hypothesis. Sentence reading times dropped an average of 175 ms for each point increase in relatedness score. The relationship was statistically significant: $F(1,743) = 32.28$, $p < 0.0001, = 0.0416$, indicating that about 4% of the variance in reading time was explained by the relatedness score. Figure 13 displays the scatter plot and regression line. The relationship held when reading times were scaled by the number of words per sentence $F(1,743) = 10.82$, $p < 0.0011$. Reading times for initial, introductory sentences (e.g. 'There was a boy named Billy.') were not used in the analysis, as their relatedness was 0. This removed one to four sentences from the beginning of each narrative.

Second, a repeated measures ANOVA was employed to test the difference in the reading times of key sentences between the high relatedness and low relatedness conditions. The ANOVA showed a significant effect, $F(1,86) = 4.78$, $p < 0.05$, indicating that reading times were slower for the low relatedness condition, confirming the hypothesis. Table 1 presents the summary statistics for each condition. The difference in conditions held when reading times were scaled by the number of words per sentence, $F(1,86) = 5.36$, $p < 0.05$.

4.1.3 *Discussion*

The results of this experiment support the MEI model of focus in narrative comprehension. As predicted by the hypothesis, sentences with higher relatedness had shorter average reading times. The regression analysis shows this general trend for the sentences in the narratives, reading times decreased an average of 175 ms for each additional point on the relatedness score. The repeated measures ANOVA shows this effect in a small set of

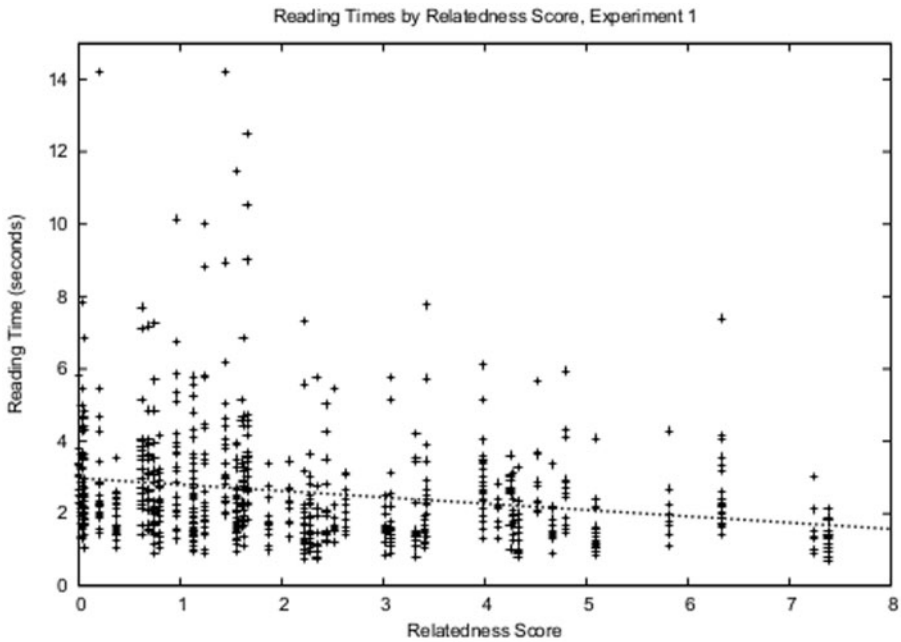

Fig. 13 Scatter plot of experiment 1 sentence reading times vs. relatedness scores with regression line, $p < 0.0001$. Sentences are read faster when they are more related

Table 1 Reading time summary statistics between conditions, experiment 1

Condition	N	Mean	Median	Std Dev	Minimum	Maximum	Skewness
High	43	2,454 ms*	1,433 ms	2,168 ms	885 ms	7,761 ms	1.944
Low	44	3,000 ms*	1,292 ms	2,677 ms	1,344 ms	7,842 ms	1.658

*$p < 0.05$.

test sentences. When the test sentences were highly related, they were read faster than in the low relatedness condition.

4.2 Experiment 2

The second experiment was designed to assess the availability of inference related information at the end of reading short narratives. This study tests online inferences without reading strategy instruction. These are inferences that are made during reading without specific instruction to do so. If the inferences are constructed by the reader, the inferences will activate related information in the reader's mental model of the story and make this information more available at the end of the narrative. On average, this activation improves a reader's recall of related words in the narrative, and it disrupts the reader's ability to recognize that other related words were not in the narrative. The overall hypothesis was that inference conditions would change the word recognition response times in favor of the inference (i.e., response would be longer if the word was not in the text and shorter if it was in the text).

4.2.1 *Method*

Materials. The narratives consisted of four filler narratives and eight experimental narratives divided

between four narratives designed to test causally necessitated inferences and four narratives designed to test intentionally necessitated inferences. The narratives were created by first creating an IPOCL story plan, selecting a Sequence of Discourse Content from the plan, and then using text templates to translate the sequence to text.

Each of the test narratives had two conditions: the prompted condition necessitated and enabled the appropriate inference according to the MEI model, and the un-prompted condition did not predict an inference due to either lack of necessitation or enablement. The conditions were the same length in sentences and approximately the same number of words per sentence. The narratives ranged from 9 to 21 sentences each, and at most 4 lines were changed between conditions. The exact narratives can be found in the Appendix of Niehaus (2009).

Procedure. The narratives were displayed one line at a time by a Java applet within a browser window. The experiment began with instructions and a short example narrative. Participants were asked to read for comprehension, and were not informed they would be timed. The participants advanced the narrative one sentence at a time. At the end of each narrative two word recognition tests were administered. The participant was presented with the question as to whether a particular word appeared in the narrative and indicated either 'yes' by pressing the 'z' key or 'no' by pressing the '/' key. In the causally necessitated inference narratives, readers were asked about a verb prompted by the inference; this word did not appear in the text. In the intentionally necessitated inference narratives, readers were asked about present in the stated intention of the character; this word did appear in the text. This difference is due to pilot findings that narratives that prompt intentional inferences with words not in the text tended to have an unusual structure, seemingly contrived for this purpose.

Design and Subjects. Eighteen students from an introductory Artificial Intelligence class participated in the study for course credit. Each participant only read one condition of each narrative. The narrative conditions were counter-balanced in a rotating sequence, and the order of the narratives given to each participant was randomized.

Table 2 Intentional narrative results, experiment 2

Condition	RT mean	RT Std Dev	% error	rd mean
Prompted	3,304 ms	2,148 ms	11.1	394 ms
Unprompted	3,179 ms	2,126 ms	9.7	401 ms

RT, response time; rd, reading time.

Table 3 Causal narrative results, experiment 2

Condition	RT mean	RT Std Dev	% error	rd mean
Prompted	4.384 ms	3.149 ms	6.9	596 ms
Unprompted	4,054 ms	3,544 ms	4.2	591 ms

RT, response time; rd, reading time.

4.2.2 *Results*

The overall results are presented in Tables 2 and 3. Reading times per word are included for completeness only. In the intentional narratives, participants were slightly faster to respond in the unprompted condition, and the error rate was slightly less in this condition as well. They were able to recall that a word relating to the intention was part of a narrative slightly faster and more accurately when the events in the narrative did not heavily involve that intention or allow for inferencing involving that intention. An ANOVA was calculated between the prompted and unprompted conditions, but no statistically significant effects were found. This result is counter to our original hypothesis.

In the causal narratives, participants were slightly faster to respond in the unprompted condition, and the error rate was slightly less. They were able to recall correctly that a word was 'not' in the narrative slight faster and more accurately when the narrative did not prompt an inference related to the word. The prompting of the inference appears to slow down response time and increase error rate. This result is in line with our original hypothesis (though without statistical significance).

Two interesting trends emerge from the data. The first is that reading times correlate significantly with response times. The second is that the variances for the response times are quite large; one standard deviation away from the mean ranges from 1 to 5 s in the intentional narratives.

Table 4 Intentional narratives, slow readers, experiment 2

Condition	RT mean	RT Std Dev
Prompted	4,484 ms	2,901 ms
Unprompted	5,763 ms	4,324 ms

Table 5 Intentional narratives, fast readers, experiment 2

Condition	RT mean	RT Std Dev
Prompted	4,285 ms*	3,462 ms
Unprompted	2,346 ms*	990 ms

RT, response time.
*$p < 0.069$, $F(1,18) = 3.53$.

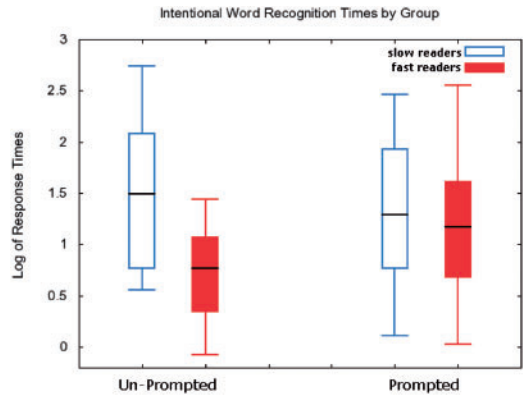

Fig. 14 Box plot of experiment 2 results

Using these observations, participants were split into two groups: slow readers (above median) and fast readers (equal or below median) based on their average reading time. The effect of the conditions on these two groups were analyzed.

Tables 4 and 5 show the results of the second analysis for the intentional narratives. Figure 14 is the box plot of the log response times. In the slow readers group, response time was faster in the prompted condition. Adding actions related to the intention seemed to improve availability, as per our hypothesis. In the fast readers group, prompting slowed the response time dramatically. Adding actions related to the intention seemed to hinder availability. This last result approached significance at $p < 0.069$.

4.2.3 *Discussion*

The experimental condition changed overall response times, providing support for the general hypothesis. In the intentionally necessitated inference narratives, the effect of the inferences depended on whether the participants were slow or fast readers. In the slow reader group, the addition of the inferences seems to make the information more available to the reader. In this case, the reader may be using the semantics of the narrative to recall the information, and is thus aided by the inferences. In the fast reader group, the prompting of intentional inferences significantly slows the response time. In this case, the participant may be relying on the surface text of the narrative more than the semantics. Reading time is faster because the reader does not have to encode the semantics of the narrative, but response time is hindered by related information. The participant takes longer to recall a specific item out of a collection of highly related items than to recall a specific item out of a collection of unrelated items.

In the causally necessitated inference narratives, the results provide weak support for the specific hypothesis. The addition of causal inferences relating to a word not in the text seems to slow response times and increase error rates. The information may be more salient in the reader's mind, and the reader may have a more difficult time discerning whether the information was inferred or included. Overall, the difference in response times over the prompted conditions supports the claim that the MEI model is able to predict online, story-level inferences.

4.3 Experiment 3

This experiment was designed to assess the ability of INFER to generate Sequences of Discourse Content that prompt inferences. The hypothesis for this experiment is that when INFER is used to prompt inferences, readers would take less time to recognize words relating to those inferences than if INFER is used to generate similar narratives that do not prompt inferences.

4.3.1 *Method*

Materials. The narratives consisted of eight filler narratives and four experimental narratives divided between two narratives designed to test causally necessitated inferences and two designed to test intentionally necessitated inferences. An extended source plan was developed to use as input to INFER. The source plan detailed the events in a western town leading up to a bank robbery. Events included purchasing a dress, a pickpocketing, drinking at the saloon, and a bank robbery. This source plan was used as the story input to INFER, and Sequences of Discourse Content were generated using single causal or intentional inference flaws, creating sequences prompting a single inference. These sequences were translated to text using simple templates, and used as the prompted condition for the experiment. Two of the four experimental narratives employed causally necessitated inferences and the other two employed intentionally necessitated inferences. The unprompted condition was constructed by switching one or more events from the sequence with steps from the story until INFER no longer predicted the inference. Two narratives were generated each with a single causal inference goal, and two narratives were generated each with a single intentional inference goal. A simple template function was used to map the narrative elements into a list of human readable sentences. The narratives ranged between five and six sentences each, and one line was changed between conditions. There was one short answer comprehension question and five word recognition questions after each narrative. The exact narratives can be found in the Appendix of Niehaus (2009).

Procedure. The narratives were displayed by a Java applet within a browser window. The experiment began with instructions and a short example narrative. In this experiment, the participants were instructed to read for comprehension and were informed they would be timed. The participants advanced the narrative one sentence at a time by pressing the spacebar.

At the end of each narrative, one comprehension question and five word recognition tests were administered. The comprehension question was a short answer question about a step in the narrative.

Table 6 Intentional narrative results, experiment 3

Condition	RT mean	RT Std Dev	% error	rd mean
Prompted	957 ms*	376 ms	6.1	2,703 ms
Unprompted	1,246 ms*	854 ms	4.1	2,818 ms

*$p < 0.05$. RT, response time; rd, reading time.

Table 7 Causal narrative results, experiment 3

Condition	RT mean	RT Std Dev	% error	rd mean
Prompted	1,097 ms	493 ms	10.2	3,333 ms
Unprompted	1,171 ms	528 ms	12.2	3,061 ms

RT, response time; rd, reading time.

The participants answered by typing on the keyboard. Next, participants were presented with the question as to whether a particular word appeared in the narrative and indicated either 'yes' by pressing the 'z' key or 'no' by pressing the '/' key.

Design and Subjects. Forty students from an undergraduate computer science course and eight students from a graduate computer science course participated in the study for course credit. Each participant only read one condition of each narrative. The narrative conditions were counter-balanced in a rotating sequence, and the order of the narratives given to a participant was randomized.

4.3.2 *Results*

The results of our study are presented in Tables 6 and 7. Figure 15 is the box plot of recognition times for the intentionally necessitated inference narratives. For the intentionally necessitated inference narratives, a two-factor mixed model ANOVA between prompted and unprompted conditions on the response times indicated a significant effect for prompted condition, $F(1,73.4) = 4.39$, $p < 0.05$, but no effect for narrative, $F(1,47.2) = 2.34$, and no interaction, $F(1,80.9) = 0.02$. The average response time was shorter when the word was prompted by an inference condition. This result indicates that the intentional inferences were constructed and that they had an effect on reading comprehension.

For the causally necessitated inference narratives, a two-factor mixed model ANOVA on the response

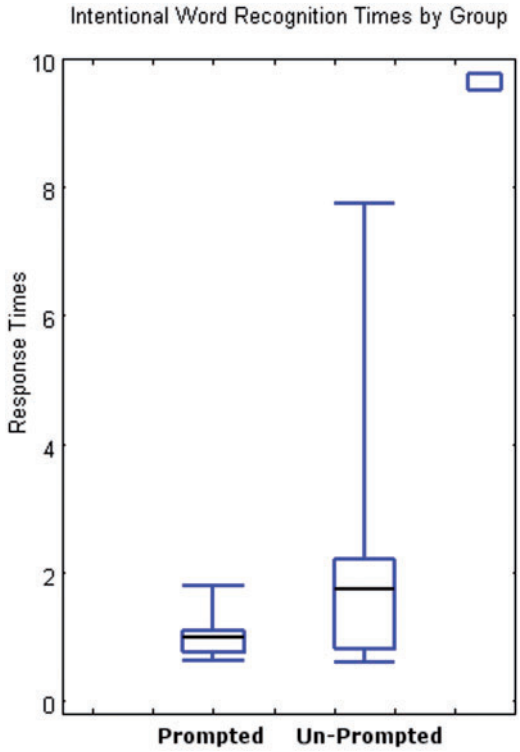

Fig. 15 Box plot of experiment 3 results. Reading times in seconds

times did not indicate a significant effect for prompted condition, $F(1,76.2) = 0.93$, but narrative was significant at $F(1,47.4) = 17.12$, with no interaction, $F(1,78.3) = 1.15$.

4.3.3 *Discussion*

These results provide support for INFER's generation of Sequences of Discourse Content to prompt inferences. Sequences generated to prompt these inferences had a significant effect on the comprehension measure of word recognition times. The intentionally necessitated inferences had a strong effect on recognition times. Recognition times decreased nearly 25% when in conjunction with a prompted inference.

The causal inferencing narratives, while decreasing response time, did not show a significant effect due to prompting. This may be because intentionally necessitated inferences are forward looking, often persisting for multiple steps in the sequence, while causal inferences are backward looking, often ending after a single step. Thus, causally necessitated inferences are enabled for shorter periods of time. For this reason their effect on comprehension may be less, making them more difficult to detect.

4.4 Discussion

The collection of these results creates a strong case for the MEI model of narrative comprehension and INFER's generation capability. The results of experiment 1 show that MEI's model of focus predicts measures of activation in readers. Participants were able to more quickly read sentences that were predicted as having higher relatedness by the MEI model. The results of experiment 2 show that MEI's model of inferences predicts changes in measures of activation of related concepts in readers. Participants in the fast reader group slowed significantly when deciding whether a word related to an inference was in the narrative (when, in fact, it was). The results of experiment 3 validated INFER's ability to generate Sequences of Discourse Content to prompt inferences. Participants were able to recognize a word in the narrative more quickly when it was related to a prompted inference. While the predictions of INFER are supported by these experimental outcomes, the authors note that the underlying representations of INFER may be completely different from how the human mind stores and manages this information; INFER only mimics the cognitive outcomes of this processing to a degree. The sum of these results indicates that INFER is operating on a valid reader model, and is able to effectively use its reader model to generate Sequences of Discourse Content with desired comprehension criteria

5 Conclusions

The principal contributions of this work are an empirically evaluated formal model of narrative focus, an empirically evaluated formal model of narrative inferencing, and an empirically evaluated method of selecting narrative discourse content to satisfy comprehension criteria. This work is unique in that it

(1) employs theory from cognitive models of narrative discourse comprehension to inform the creation of computational models of narrative comprehension and (2) employs these models for generation. The work defines an explicit computational model of a reader's comprehension process during reading, predicting aspects of narrative focus and inferencing. Focus is derived from activation values in an association network. Inferences about the story world are generated by a partial-order planner. This reader model is employed in a narrative discourse generation system to select content from a source plan, creating discourses that satisfy comprehension criteria. The generation component defines a novel partial-order planning algorithm to satisfy the criteria. This work has implications for both cognitive scientists and builders of narrative generation systems. For cognitive scientists, new, more precise models of narrative comprehension may be defined and tested, removing possible researcher bias and improving the independence of experimental measures. For implementers of systems, the ability to generate narrative discourses that are more coherent and cohesive may improve the effectiveness of narrative systems in entertainment, training, and education. The ability to predict aspects of comprehension in generated discourses gives the implementers new power for influencing narrative experiences.

Although this work is unique in its approach, there are some limitations. First, although the computational complexity of the planning algorithm matches the complexity of the decision task presented by the problem, the complexity is still generally nonpolynomial. Second, the specification of narrative content solely by the comprehension criteria is not natural to authors. This relates to the third limitation, the expressiveness of the reader model is limited to two main properties of comprehension, while cognitive models of narrative discourse comprehension have found evidence for many differing properties of comprehension. Fourth, while the empirical evaluation of the presented model covered the three main contributions, the empirical evaluation of narrative generation systems is complex and difficult, and more experiments may provide further evidence for or against components of the system. Lastly, this algorithm requires representation of the story world in an IPOCL plan structure, including the logical preconditions and effects of steps, causal relationships, and temporal relationships. This representation may not always be convenient for nonplanning-based systems, and other representations that include characters, causality, and temporality may be explored in future systems.

There are several possible areas of future work. The reader model may be extended to account for more aspects of narrative comprehension: other forms of inference, the skill of the reader, situated reasoning models. The comprehension criteria may be modified to provide a more natural means of expressing authorial control over the reading experience, and the new requirements may be propagated back into the model. Finally, generating sequences of discourse content from a planning domain, rather than a source plan, is an intriguing prospect. In such a system, the planner is responsible not only for choosing content at the discourse level, but also constructing a coherent story level to adhere to comprehension criteria.

References

Allbritton, D. (2004). Strategic production of predictive inferences during comprehension. *Discourse Processes*, **38**(3): 309–22.

Bae, B. C. and Young, R. M. (2008). A use of flashback and foreshadowing for surprise arousal in narrative using a plan-based approach. In *Proceedings of the International Conference on Interactive Digital Storytelling (ICIDS 08)*, Erfurt, Germany, November 2008.

Casteel, M. A. (2007). Contextual support and predictive inferences: What do readers generate and keep available for use. *Discourse Processes*, **44**(1): 51–72.

Cheong, Y. G. (2007). *A Computational Model of Narrative for Suspense*. Ph.D. thesis, North Carolina State University.

Cheong, Y. G. and Young, R. M. (2008). Narrative generation for suspense: Modeling and evaluation. *In Proceedings of the International Conference on Interactive Digital Storytelling (ICIDS 08)*, Erfurt, Germany, November 2008.

Cullingford, R. (1977). Sam: A program that uses world knowledge to understand. *ACM SIGART Bulletin*, **61**: 53–4.

DeJong, G. (1979). Prediction and substantiation: A new approach to natural language processing. *Cognitive Science*, **3**(3): 251–73.

Frank, S. L., Koppenc, M., Noordmana, L. G. M., and Vonk, W. (2003). Modeling knowledge-based inferences in story comprehension. *Cognitive Science*, **27**: 875–910.

Gerrig, R. (1993). *Experiencing narrative worlds on the psychological activities of reading*. New Haven, CT: Yale University Press.

Graesser, A. and Clark, L. (1985). *Structures and procedures of implicit knowledge*. Norwood, NJ: Ablex.

Graesser, A., Singer, M., and Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, **101**(3): 371–95.

Graesser, A., Olde, B., and Klettke, B. (2002). How does the mind construct and represent stories? In Green, M. C., Strange, J. J., and Brock, T. C. (eds), *Narrative Impact: Social and Cognitive Foundations*. New York, NY: L. Erlbaum Associates.

Hobbs, J. R., Stickel, M., Martin, P., and Edwards, D. (1988). Interpretation as abduction. In *Proceedings of the 26th annual meeting on Association for Computational Linguistics*, Morristown, NJ, USA: Association for Computational Linguistics, pp. 95–103.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction integration model. *Psychological Review*, **95**: 163–82.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, MA: Cambridge University Press.

Langston, M. C., Trabasso, T., and Magliano, J. P. (1999). A connectionist model of narrative comprehension. In *Understanding Language Understanding*. MIT Press.

Lebowitz, M. (1984). Creating characters in a story-telling universe. *Poetics*, **13**: 171–94.

Lebowitz, M. (1985). Story-telling as planning and learning. *Poetics*, **14.6**: 483–502.

Linderholm, T. (2002). Predictive inference generation as a function of working memory capacity and causal text constraints. *Discourse Processes*, **34**: 259–80.

Magliano, J. P., Miller, J., and Zwaan, R. A. (2001). Indexing space and time in film understanding. *Applied Cognitive Psychology*, **15**: 533–45.

Magliano, J. P., Taylor, H. A., and Kim, H-J. J. (2005). When goals collide: Monitoring the goals of multiple characters. *Memory and Cognition*, **33**(8): 1357–67.

Mateas, M. and Stern, A. (2003). Façade: An experiment in building a fully-realized interactive drama. In *Game Developers Conference, Game Design track*, March 2003. http://www.interactivestory.net.

McKoon, G. and Ratcliff, R. (1992). Inference during reading. *Psychological Review*, **99**(3): 440–66.

Mo, L., Chen, H., Li, Y., Chen, Z., and He, X. (2007). Effects of event-related centrality on concept accessibility. *Discourse Processes*, **43**(3), 2007.

McAllester, D. and Rosenblatt, D. (1991). *Systematic nonlinear planning*. MIT AI Memo No. 1339. 1991.

Mott, B. and Lester, J. (2006). U-director: A decision-theoretic narrative planning architecture for storytelling environments. In *Proceedings of the Fifth International Conference on Autonomous Agents and Multiagent Systems*, Hakodate, Japan, pp. 977–84.

Mueller, E. T. (2002). Story understanding. *Encyclopedia of Cognitive Science*, vol. 4. London: Nature Publishing Group, pp. 238–46.

Mueller, E. T. (2003). Story understanding through multi-representation model construction. In *Text Meaning: Proceedings of the HLT-NAACL 2003 Workshop*.

Mueller, E. T. (2007). Understanding goal-based stories through model finding and planning. In Magerko, B. and Riedl, M. (eds), *Intelligent Narrative Technologies: Papers from the AAAI Fall Symposium*. Menlo Park, CA: AAAI Press, pp. 95–101.

Niehaus, J. (2009). *Cognitive models of discourse comprehension for narrative generation*. NC, USA: North Carolina State University.

Norvig, P. (1987a). Inference in text understanding. In *Proceedings of the National Conference on Artificial Intelligence*, pp. 561–65.

Norvig, P. (1987b). *A Unified Theory of Inference for Text Understanding*. Ph.D. thesis, University of California, Berkeley.

Penberthy, J. S. and Weld, D. (1992). UCPOP: A sound, complete, partial order planner for adl. In *Proceedings Third International Conference on Principles of Knowledge Representation and Reasoning (KR-92)*, pp. 103–14.

Ram, A. and Moorman, M. (eds) (1999). *Understanding Language Understanding: Computional Models of Reading*. MIT Press.

Riedl, M. O. (2004). *Narrative Planning: Balancing Plot and Character*. Ph.D. thesis, North Carolina State University.

Riedl, M. O. and Sugandh, N. (2008). Story planning with vignettes: Toward overcoming the content production bottleneck. In *Proceedings of the 1st Joint International Conference on Interactive Digital Storytelling (Formerly TIDSE and ICVS)*, Erfurt, Germany.

Riedl, M. O. and Young, R. M. (2004). An intent-driven planner for multi-agent story generation. In *Proceedings of the 3rd International Joint Conference on Autonomous Agents and Multi Agent Systems, New York*.

Riedl, M. O. and Young, R. M. (2005). Story planning as exploratory creativity: Techniques for expanding the narrative search space. In *Proceedings of the 2005 IJCAI Workshop on Computational Creativity*, Edinburgh.

Rinck, M. (2000). Temporal and spatial distance in situation models. *Memory and Cognition*, **28**: 1310–20.

Schank, R. (1975). Sam: A story understander. Technical Report 46, Yale University, CT, USA.

Sgouros, N. (1999). Dynamic generation, management and resolution of interactive plots. *Artificial Intelligence*, **107**: 29–62.

Shapiro, S. C. and Rapaport, W. J. (1995). An introduction to a computational reader of narratives. In Duchan, J. F., Bruder, G. A., and Hewitt, L. E. (eds), *Deixis in Narrative: A Cognitive Science Perspective*. Hillsdale, NJ: Lawrence Erlbaum, pp. 79–105.

Shears, C., Miller, V., Ball, M., Hawkins, A., Griggs, J., and Varner, A. (2007). Cognitive demand differences in causal inferences: Characters' plans are more difficult to comprehend than physical causation. *Discourse Processes*, **43**(3): 255–78.

Szilas, N. (2003). Idtension: A narrative engine for interactive drama. In *Proceedings of the 1st International Conference on Technologies for Interactive Digital Storytelling and Entertainment*. Darmstadt, Germany.

Szilas, N. (2007). A computational model of an intelligent narrator for interactive narratives. *Applied Artificial Intelligence*, **21**(8): 753–801.

Therriault, D. J. and Raney, G. E. (2007). Further exploration of time in situation models. evidence for the representation of duration information. *Discourse Processes*, **43**(2): 173–200.

Trabasso, T. and Wiley, J. (2005). Goal plans of action and inferences during comprehension of narratives. *Discourse Processes*, **39**(2): 129–64.

Turner, S. R. (1992). *MINSTREL: A Computer Model of Creativity and Storytelling*. Ph.D. thesis, University of California, Los Angeles.

Wilensky, R. (1976). Using plans to understand natural language. In *ACM 76: Proceedings of the annual conference*. New York, NY, USA: ACM Press, pp. 46–50. doi: http://www.lib.ncsu.edu:2268/10.1145/800191.805524.

Young, R. M., Riedl, M. O., Branly, M., Martin, R. J., and Saretto, C. J. (2004). An architecture for integrating plan-based behavior generation with interactive game environments. *Journal of Game Development*, **1**: 20.

Zwaan, R. A. and Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, **123**(2): 162.

Zwaan, R. A., Langston, M. C., and Graesser, A. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science*, **6**(5): 292–97.