Using Virtual Narratives to Explore Children's Story Understanding

Paper id #123

ABSTRACT

Interactive Narratives are systems that use automated narrative generation techniques to create multiple story variants which can be shown to an audience, as *virtual* narratives, using cinematic staging techniques. Previous research in this area has focused on assessment of aspects such as the quality of the automatically generated narratives and their acceptance by the audience. However in our work we deviate from this to explore the use of interactive narratives to support cognitive psychology experiments in story understanding. We hypothesized that the use of virtual narratives would enable narrative comprehension to be studied independently of linguistic phenomena. To assess this we developed a demonstration interactive narrative featuring a virtual environment (Unity3D engine) based on a pre-existing children's story which allows for the generation of variants of the original story that can be "told" via visualization in the 3D world. In the paper we introduce a narrative generation mechanism that provides control over insertion of cues facilitating story understanding, whilst also ensuring that the plot itself is unaffected. An intuitive user interface allows experimenters to insert and order cues and specific events while the narrative generation techniques ensure these requests are effected in a consistent fashion. We also report the results of a field experiment with children (age 9-10) that demonstrates the potential for the use of virtual narratives in story understanding experiments. Our results demonstrated acceptance of virtual narratives, the usability of the system and the impact of cue insertion on inference and story understanding.

Keywords

Virtual Agents; Interactive Storytelling; Narrative Modeling; Planning; Game-based Education.

1. INTRODUCTION

Interactive Narrative systems (IN) which feature automated narrative generation have been shown capable of creating multiple story variants, both in terms of plot evolution and discourse-level presentation, using cinematic staging techniques in 3D virtual environments [3, 8]. To date the main target application of such systems has been entertainment [18] as well as training [24] and education [21]. Some previous research has investigated cognitive aspects of narrative understanding in order to assess the quality of automatically generated narratives, such as whether system generated narratives were understood by users. However, as far as we are aware, no work has explored the use of IN to support cognitive psychology experiments in story understanding – an important research topic in developmental psychology. Here, the use of IN offers tremendous potential through their ability to tell "virtual stories" by offering the potential to expand the study of narrative comprehension so that it is independent of linguistic phenomena and linguistic abilities and instead relates purely to narrative phenomena. Furthermore, this allows possible long-term applications such as narrative comprehension experiments with children younger than 6 with little or no reading abilities [10]. Another strength of IN systems, and automated narrative generation, is the possibility of generating story variants where specific determinants of story understanding are controlled from first principles, thus allowing for various hypotheses to be tested. This is important as theories of narrative understanding require a reader to construct a coherent mental representation of the text [6, 11, 27]. The reader needs to establish meaningful relations between concepts, facts, intended meaning, and ideas from the text with their background knowledge, and integrate all information in a situation model. This implies that readers need to monitor several story elements, such as the story setting, the events, actions taking place, and the protagonist and their motives and goals [23, 30]. Moreover, it has been established that narrative comprehension skills tend to transfer across different media [9].

Hence our motivation was to explore the application of IN techniques in this area by addressing the following challenge: how to develop a narrative generation mechanism which would enable the production of story variants in which specific determinants of story understanding could be controlled from first principles, and various hypotheses could be tested, for instance on the role of in-story cues in causal understanding; supported by the visualization story variants within a 3D virtual world using cinematic staging techniques to support in-story cues. To this end we have developed a mechanism that allows control over the high-level specification of narrative variants during the design of story understanding experiments and which are used during narrative generation.

In order to demonstrate the potential of this approach we have implemented it within a demonstrator IN featuring a 3D storyworld based on a children's story called "The day Tuk became a hunter" [13]: a story about an Inuit boy who lives in the Arctic and dreams of becoming a great hunter. This system was used in a series of experiments which we report, and which demonstrated the potential extent of the inclusion of narrative cues through generative mechanisms allowing the evaluation of story understanding for various narrative instances.

The paper is organised as follows: section 2 considers related work and section 3 discusses issues relating to the modelling of narrative for story understanding experiments. In section 4 we

Appears in: *Proceedings of the 16th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2017), S. Das, E. Durfee, K. Larson, M. Winikoff (eds.), May 8–12, 2017, São Paulo, Brazil.*

Tuk is an Inuit boy who lives with his family in an Igloo in the Arctic. He dreams of becoming a great hunter like his father. Eventually he is allowed to join his father on a hunt and they set off. The father attempts to hunt a Seal but is attacked and disarmed by a Polar Bear. Tuk rescues his father and then he is then in turn rescued by his father, following which they both flee back to their Igloo. However, the Polar Bear follows them and they are forced to hide inside the Igloo and wait for the Polar Bear to leave. Eventually the family run out of food and the father announces that they must kill the Polar Bear or die. That night, Tuk sneaks out of the Igloo while the Polar Bear is asleep and carves a large polar bear out of snow. In the morning the real Polar Bear awakes and, seeing the snow Polar Bear, is scared and runs away. The story ends with Tuk's father seeing the Polar Bear fleeing, hugging his son and declaring him to be 'Tuk the Hunter'.

Figure 1: Temporal order of events in the Story of Tuk. There are several themes: Tuk's goal of becoming a hunter and recognition from his father, dangers of life in the Arctic, importance of family and Tuk's use of his carving ability. It is Tuk's carving ability that is key to understanding how he defeats the polar bear. The linear nature of the story facilitates its formalization as a Planning domain.

outline our narrative control mechanism for generation of variants for story understanding experiments. In section 5 we overview our implemented demonstrator IN system and in section 6, give a detailed example of the generation of narrative variants. System generated narrative variants were used in a series of experiments which we report in section 7 and we conclude in section 8.

2. RELATED WORK

Previous research in IN has considered cognitive aspects, in particular narrative understanding by users, to assess the quality of narrative generation [3]. For instance, some researchers have adapted narrative understanding methodologies, such as the QUEST questionnaire, originally developed to evaluate causal understanding of written stories [5], to that of interactive narratives [19], in particular causal understanding and inference [20]. The rationale for this approach was to evaluate whether automatic narrative generation indeed preserved the type of causal relationships between events, known to be central to narrative understanding. Interestingly, the QUEST methodology has also been used to evaluate the impact of camera placement and real-time editing on narrative understanding [8], confirming the role of discourse-level presentation in the causal understanding of narratives.

Other research has used virtual worlds and 3D animations to explore the perception of causality in physical phenomena [28]: however, this should not be confused with the more conceptual causal understanding at play in narratives. In other work, the focus has been maintenance of story consistencies across multiple interactive narrative story paths [22], or even the study of the effects of technology-enhanced stories for young children's literacy development when compared to listening to stories in more traditional settings like storybook reading [25].

There has been significant previous work in developing INs for children, and for most of them the purpose has been educational rather than recreational. The Fearnot! system [1, 26] used INs to educate school children on bullying issues, by fostering empathy towards virtual agents. Follow-up research in the MIXER system [2] extended this work to the understanding of multiculturalism. While both these systems were essentially leveraging emotional responses to promote desirable behavior, other INs have been centered on conceptual learning. More specifically, the Crystal Island [15] system developed an interactive narrative embedding the 8th-grade microbiology curriculum and was used to explore learning benefits [12], as well as supporting experiments on users' goal recognition [14] and customization of narrative for pedagogical drama [16].

Compared with previous work, the research we present here is specifically concerned with basic phenomena in narrative understanding: narratives are thus the experimental medium as well as the object of study, rather than domain knowledge or affective responses. Unlike previous interactive narrative work targeting adult users, in which causal understanding was explored to measure the quality of narrative generation [3, 19] or presentation [8], here narrative generation is specifically used to explore causal understanding itself.

3. NARRATIVE REQUIREMENTS

Our overall motivation in this work was to explore the use of INs to support experiments in story understanding. Central to this was the ability of the system to generate different "visual story" variants to be used in experiments and which provided support for control of the positioning and visualization of narrative cues during story generation without altering the plot itself. This is critical as the arrangement and presentation of cues will, for instance, influence causal inferences within story understanding. Hence the requirement was to develop a mechanism that enables manipulation of those aspects of the story that affect generation of causal inferences by participants, while freeing experimenters from issues relating to the practical aspects of generating suitable narrative variants – and in a way which is supported by narrative visualization (ie how the story is shown to viewers). In other words, the system is tasked with maintaining story consistency and taking care of low-level plot details and the experimenter takes high-level control over the details of the story that is presented to participants.

There is no shortage of narrative models and theories, however in this work we adopt a Psychological framework to identify precise requirements for computational narrative generation in this context. According to Stein and Glenn [23], the structure of a story can be divided in two parts: the Setting category and all Episodes. The protagonist and the spatial and temporal surroundings in which the story takes place are introduced in the Setting of a story. The information presented in the setting allows readers to interpret information and all subsequent events. The Episode is a higher order unit of analysis in a story and contains a sequence of five different categories, namely the Initiating Event, Internal Response, Attempt, Consequence and Reaction. Each category contains specific types of information and serves a different function in the schema.

Of particular interest in the context of narrative generation for story understanding are protagonist *Goals*, part of the Internal Response category, and *Actions* which are part of Attempt. Goals are seen as the most critical part of the story since story knowledge is organized around them. Goals also provide the motivation for activities carried out by the protagonist. Actions are those activities carried out by the protagonist in pursuit of their goals.

Thus we hypothesized that emphasizing the protagonists goals along with motivation and actions would improve story understanding, as these types of story "grammar" are closely related to the causal structure of a narrative. We refer collectively to these elements as *narrative cues*. This was incorporated into our narrative framework via the following mechanisms (also shown in Fig. 2):

- A Narrative Generator featuring a mechanism to control the se-

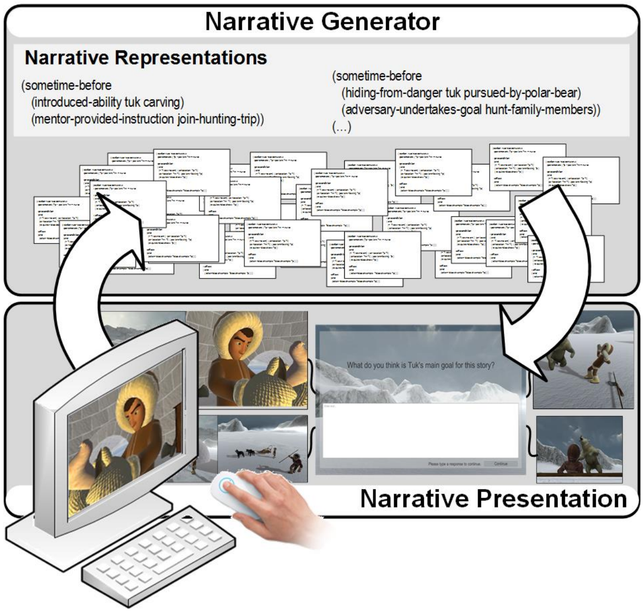


Figure 2: Plan-based narrative generation which allows user control over the narrative variants (via interface), with real-time 3D visualization that permits punctual evaluation of story understanding.

lection (insertion and positioning) of story content related to the protagonists goals and actions i.e. the narrative cues (section 3).

- The use of Narrative Presentation (Cinematic Staging) Techniques to support the presentation of selected narrative cues.

3.1 Narrative Cue Specification

Narrative cues are derived from existing elements of the narrative model. For story understanding experiments and manipulation of goals, actions and motivations there is a requirement to be able to: include or exclude selected narrative cues; specify a particular manner of presentation within the version of the story presented to a participant; and specify the temporal order of selected narrative cues. In our solution, narrative cues are specified by the user (experimental designer) using the graphical user interface (GUI) (see Fig. 2 & 6).

An example of narrative cue selection from the story of Tuk, where several cues can function to convey the danger of the situation faced by Tuk and his family while being hunted by the Polar Bear: one cue controls whether characters provide warnings to each other; another cue reinforces, through camera close-ups, the diminishing supply of food; and a third cue reinforces, again through camera close-ups, the ferociousness of the Polar Bear. Each of these cues can be controlled individually and employed (or ignored) for a particular effect on the presented story.

It is also possible to exercise temporal control over cues by adjusting the order that story events are presented to participants. An example of this is the timing of the introduction of Tuk's ability at carving animals which plays an important role in the resolution of the story towards the goal. This scene can be included, as in the original story, near the beginning before Tuk's encounter with the Polar Bear or moved to a later point, such as when the family are trapped in their igloo, so that it is closer to the resolution. This example featured in our experiments: please refer to section 6.

Figure 3: Key scenes from the part of the story where the Polar Bear is frightened away by the Snow Bear that Tuk has carved. This is an important narrative cue in those variants of the narrative when the actual work of Tuk carving the Snow Bear is not shown.

4. NARRATIVE PRINCIPLES

Our implemented system features an AI Planning-based narrative generator which is able to generate narratives that strike a balance between introduction of cues or minor presentation changes and the preservation of the overall plot structure. In line with plan-based narrative generation, the insertion of cues and required plot structure can be encoded as part of the the narrative planning "problem" – specified in terms of an initial state of the world augmented with a series of constraints specifying required properties to be realized in the output narrative, along with a goal condition; and where the narrative itself is a sequence of pre- and post-condition actions that map the initial state into some state in which the goal condition is true. Below we detail the features of the narrative problem, actions and generation method that ensure narrative variants conform to experimenter specification (see Fig. 6).

4.1 Narrative Planning Problem

Within our approach, the experimenter specified narrative cues form part of the input to the narrative generator. Also part of the narrative generator input is the goal condition which, given the need to strictly enforce the arc of the baseline story, (in our worked example, the story of Tuk), always remains the same. From this input specification and goal, a narrative planning problem is automatically generated, modelled using PDDL3.0.

As an example, for the story of Tuk the goal of Tuk being recognised as a great hunter, is represented as:

(**at-end** (recognised-brave-hunter tuk))

Any ordering of story content required by the baseline story arc is represented using the *sometime-before*[1] operator which are included in the problem file [4]. For example, the constraint;

(**sometime-before**
(recognise-as-brave-hunter father tuk)
(frighten-away tuk polar-bear snow-bear))

captures the ordering that the polar bear must be frightened off and run away **before** Tuk can be recognized as a brave hunter.

In addition to such fixed ordering of events, all narrative cues selected and ordered by the experimenter at the interface are also mapped to constraints in the narrative planning problem, again represented using the *sometime-before* modal operator. In total, the Story of Tuk supports 14 specific narrative cues, mapping into the following categories based on the causal inference they support:

- inferences about the danger of the situation faced by the family
- inferences about Tuk's carving ability

[1]The semantics of *(sometime-before A B)* requires that application of actions in solution plans make B true **before** A.

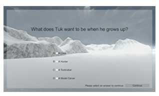
①

```
(:action introduce-protagonist-ability
  :parameters (?c - character ?a - ability)
  :precondition (and
     (ability ?c ?a)
     (= (reinforced-ability ?a) 1) ... )
  :effect (and
     (increase (reinforced-ability ?a) 1) ... ))
```

②

```
(:action recognise-bravery
         :parameters (?t ?f ?p - character)
  :precondition (and
     ((adversary ?p) (protagonist ?t)
     ((mentor ?t ?f) (frightened-away ?t ?p) )
  :effect (and
     ((acknowledged-bravery ?f ?t) ... ))
```

③



```
(:action ask-questionnaire
  :parameters (?q - questionnaire)
  :precondition (and
     (not (asked-questionnaire ?q )))
  :effect (and
     (asked-questionnaire ?q )))
```

Figure 4: Example narrative actions formalized as PDDL operators: (1) the introduce-ability action is one of several that reinforce Tuk's carving ability (increment fluent value giving indication of strength); (2) Tuk's goal condition is achieved by the action of his father recognizing his bravery; (3) dedicated questionnaire actions are generated as part of the narrative for integration into the visualization.

- inferences concerning Tuk's desire to be like his father and receive recognition as a hunter.

From an implementation perspective, the system uses the constraints to generate a bespoke PDDL3.0 problem file to be used to generate an output narrative. Hence, when an experimenter opts to select (or not) a narrative cue this ordering constraint is included (omitted) from the specification of the narrative planning problem. This problem then captures both the fixed story structure and the various properties (specific cues) to be realized by the generated plan.

4.2 Narrative Planning Actions

Narrative planning actions are modelled as pre- and post-condition actions with each providing a generic description of narrative events that move the story forward in some way, i.e. change the state of the narrative world and the situation of virtual characters within it.

For the story of Tuk, the domain model consists of a total of 100 generic narrative actions which yield in excess of 300 instantiated actions at run time, with generated narrative plans typically containing 65 to 75 narrative actions, depending on the specific cues selected. Run-time for visualization of full narratives is upwards of 15 minutes. For illustration, some example PDDL actions from the story of Tuk domain model are shown in Fig. 4.

4.3 Narrative Generation

The narrative generator in the system has been developed using inspiration from the decomposition narrative planning approach of Porteous et al. [17]. Their approach is useful in the context of our work since it provides a mechanism to control the "shape" of generated narratives through the use of authored narrative constraints which represent partial descriptions of states of the narrative world that are desired in the narrative. Partially ordered authored constraints are linearised and used as intermediate goals to incrementally generate a narrative. The advantage of this approach is that it allows our system both to enforce the required structure of the Tuk story arc and also experimenter specified variations thus freeing the experimenter from concern with dependencies which may not be directly specified by the causal cue. At a basic level, this means that the experimenter can select the inclusion of specific causal cues, without having to know the various points throughout the story where the changes will be realised. This is important since it allows for experimenters to abstract content in their specification, taking advantage of the fact that it will be automatically incorporated at narrative generation time, should it be required as part of plan consistency.

For example, in the original telling of the story of Tuk, his carving skills are introduced at the end of the day so this is preceded by a sunset scene to show the transition to evening. If the carving scene is moved to follow a scene where it is already evening, then no transition sunset scene is needed and this is handled automatically.

Baseline Narrative Problem	**Alternate variant Problem**
(*sometime-before*	......
(introduced-ability tuk carving) ①	(*sometime-before*
(invited-hunting-trip tuk father)) ;b1	(introduced-ability tuk carving) ①
(*sometime-before*	(wait-for-end-danger)) ②
(task-completed pack-dog-sled) ;b2	(*sometime-before*
(introduced-ability tuk carving)) ①	(request-help father tuk) ;a1
......	(introduced-ability tuk carving)) ①
(*sometime-before*	(*sometime-before*
(wait-for-end-danger) ②	(contemplate-carving-plan tuk) ③
(...)) ;bn	(request-help father tuk)) ;a1
Order of Constraints	Order of Constraints
b1 → ① → b2 → .... → bn → ②	... →② → ①→ a1 → ③

Figure 5: Examples of the use of PDDL *sometime-before* constraints to enforce the order of presentation of Tuk's carving ability cues (selected and ordered by the experimenter at the GUI, Fig. 6): the baseline problem specifies the carving cue to occur early in the narrative (before Tuk departs on the hunting trip); the alternative has it at a later stage (after Tuk and his family have waited for the danger from the polar bear to end and his father asks for help). Both narratives are illustrated in Fig. 8.

5. SYSTEM OVERVIEW

As already mentioned, in order to demonstrate our IN approach to the study of narrative understanding we have developed a prototype system implementing our narrative generation control mechanism and our approach to animation of virtual narratives. This system was subsequently used in a series of field experiments. The system is shown in Fig. 6 and consists of three central components: an interface for configuration of narrative understanding experiments ①; a plan-based narrative generator ②; and a 3D real-time visualization system developed with the Unity3D game engine ③.

The interface makes it possible for experimenters to enter a high-level specification of the properties of narrative variants that are to be generated. More specifically these properties translate to an experimental hypothesis (based on the role of narrative cues, distance between cues and character activities that are related to them, and the role of intervening activities in suppressing inference) which is translated by the system into control criteria which are in turn used to generate the narrative variant. This offers control to the experimenter who is freed from having to directly manipulate the narrative generation mechanism itself. This high-level specification also abstracts from the story presentation details, which are embedded in

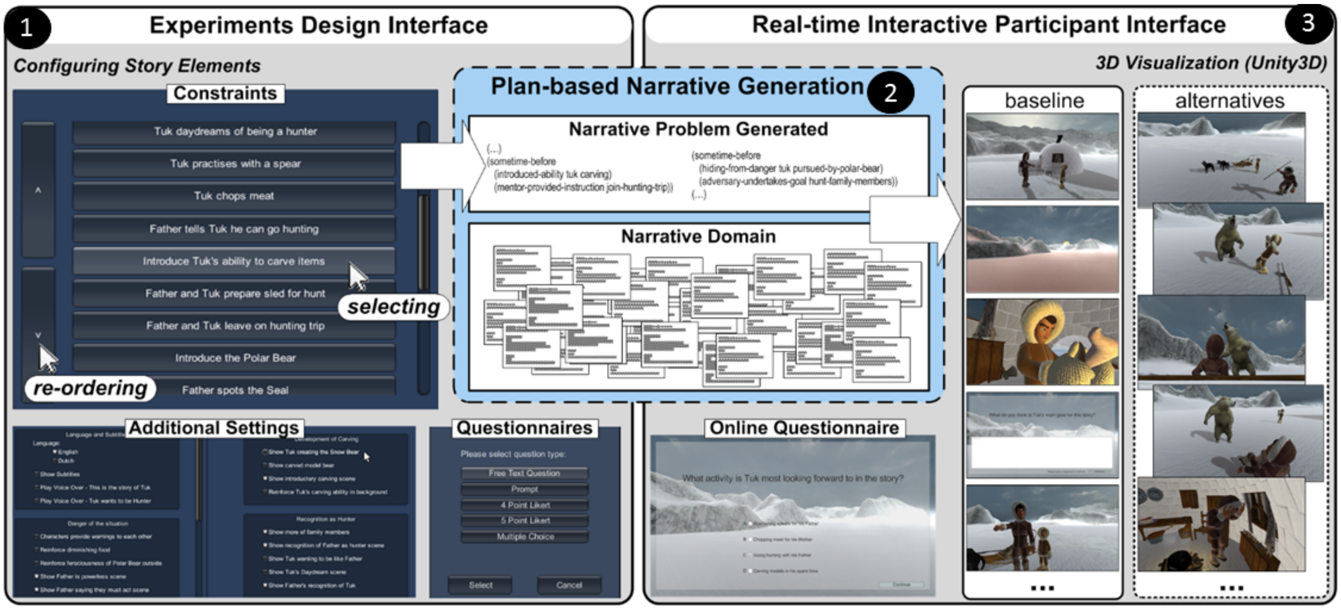


Figure 6: System Overview. (1) Experimental Design Interface: select and order story cues; design questionnaires. (2) Generate story variants. (3) During experiments: system generated narratives resented to participants as 3D visualization using the Unity3D game engine. Elicitation and collection of online questionnaires is integrated with the narrative presentation. See section 5 for further detail.

narrative generation itself (as shown in Fig 6). Hence emphasizing a specific cue does not require additional editing or modification of camera shots: instead low-level mechanisms for emphasizing an action are produced automatically during narrative generation. The mechanism used for narrative configuration is detailed in section 3. Experimenters can also use the interface to set up questionnaires to be used to collect responses from participants. A further advantage of using the GUI is that it enables the experimenter to specify narrative parameters that will be used to generate the narrative variant without any need for them to be familiar with the actual narrative generation technique used as it only deals with simple operations (such as inclusion, re-ordering) on selected elements of the story content (as discussed later in section 3).

At run-time the baseline narrative and system generated narrative variants are visualized in the 3D story world, using a mapping of narrative actions to animation scripts for the Unity3D game engine which are instantiated and can be viewed by story understanding experiment participants (for an impression of the visuals, see Fig 6 and Fig 3).

The animation follows the text version of the story variant as closely as possible, and is essentially a non-text visualization of the story, with any text and captions removed as much as possible in the animated version. This is done in order to rule out a confound with (technical) reading abilities of the participants, and also to explore pure narrative understanding phenomena that only depends on story structure and action presentation. Where no reasonable visualization of the text is possible a narration is used (e.g. the first sentence of the story of Tuk, which states "*This is a story about a boy named Tuk, who lived in the Arctic*", was narrated, as no visualization of this setting information was possible).

6. INTERACTIVE NARRATIVE EXAMPLE

As an illustration, we consider how the baseline story of Tuk was generated and the variants used in our experiment (detailed in section 7.1). The experiment concentrates on the role of Tuk's carving skills in the resolution of the story of Tuk and the impact of when this is introduced and reinforced on children's understanding of the story. In the baseline (original) story (shown on the left hand side of Fig. 8), Tuk's carving ability is introduced early on (action number 20). Whereas for the alternate variant (right hand side of Fig. 8), which aimed to aid the children's understanding of the inference related to Tuk's ability, the carving skill was introduced later, closer to the narrative action making use of that skill, and reinforced by a subsequent action (actions 50 and 61). It is important to note that Tuk's skill at carving plays a central role in the understanding of the story – Tuk makes cunning use of it to carve a snow bear to frighten away the polar bear – and since the system makes it possible to remove the scene where Tuk is seen carving the bear, it thus creates the need for an inference from the participant.

As illustration, part of these problem files are shown in Fig. 5. On the left hand side are a selection of constraints which encode the user requirements. The semantics of these orders are that *(sometime-before A B)* requires that *B* is made true in the narrative at some point before *A* becomes true. Consequently they support the ability to introduce cues in various contexts and at differing distances from the inferences required for the experiments. For example, here they capture the requirement that introduction of Tuk's carving ability ① is ordered to occur **before** waiting for the danger to end ②. Down the right hand side are shown part of the encoding for the alternate narrative which has the introduction of Tuk's carving ability occurring much later in the narrative: the constraints force this to not occur until **after** the waiting phase ②.

PDDL constraints were also used to control the position in which participants are asked questions about the narrative. The questions used in our experiments are summarized in Fig. 7. Relevant to this example is the question relating to Tuk's carving ability, question B(3). Its position in the narrative is enforced, in the baseline and alternate narrative, via the constraint below (and shown in Fig. 8):

(**sometime-before**
  (asked-questionnaire questionnaire-b)
  (contemplate-carving-plan tuk))

Narrative generation proceeds by linearizing the specification of

Part (A): questions related to Tuk's goals and desires

1: What do you think is Tuk's main goal for this story?
2: What activity is Tuk most looking forward to in the story?
3: What does Tuk want to be when he grows up?
4: Who does Tuk want to be most like when he grows up?

Part (B): questions related to the family in danger

1: How do you think Tuk might defeat the Polar Bear?
2: The family are too scared to leave the igloo because of the Polar Bear. Why do the family most want to get rid of the Polar Bear?
3: What special skill does Tuk have?

Part (C): questions related to the resolution of the story.

1: What was Tuk's main goal in this story?
2: Did Tuk achieve his main goal?
3: Why did the Polar Bear run away? Please explain the reasons as fully as possible.
4: Who was responsible for scaring away the Polar Bear?
5: What was the Polar Bear scared of?
6: How was the snow bear created?
7: What did you think of the story?
8: How much would you agree with the following statement? I found the story interesting.
9: How much would you agree with the following statement? I found the story confusing.
10: How much would you agree with the following statement? I found the story entertaining
11: How much would you agree with the following statement? I enjoyed watching the story.
12: How much would you agree with the following statement? I would recommend the story to my friends.

Figure 7: Summary of Questions used in Experiments: (A) deals with children's understanding of Tuk's goals and motivations; (B) are related to the family being in danger trapped in the igloo and how this situation might be resolved; (C) are questions related to the resolution of the story and the overall experience. Part (C) is administered after completion of the story.

sometime-before constraints (as shown in Fig. 5). The narrative is then generated incrementally starting from the initial state of the narrative world and taking the next node in the order as a goal for generation of the next phase of the narrative. A sub-narrative is generated for each such phase of the narrative using the classical planner METRIC-FF [7], the state of the narrative world is advanced (updated using the pre-conditions of the actions in the narrative). This continues until all nodes in the order have been considered.

This approach is well suited to the application since its high level control mechanism is easily configurable to experimental settings and automatically manages narrative consistency hence relieving experimenters of the burden of low level detail. Fig. 8 gives an overview of the resulting narratives and an impression of the visualization via thumbnails.

7. EVALUATION

To assess the potential of INs for story understanding experiments we staged a series of experiments using our implemented system with school children as participants. The aim was to evaluate the ability of the system to influence story comprehension as well as the overall acceptability of the virtual stories and their automatic variants, from the children's perspective. The latter aspect should ensure that generated narratives are not perceived as experiments or tests but retain the engaging properties of virtual stories. This is also important to preserve ecological conditions of story understanding.

7.1 Experiment

We conducted our experiments in a primary school, recruiting 53 children who were all native English speakers. We followed approved procedures at our Institution and all experiments were supervised onsite by school staff, obtaining individual parental consent and preserving children's anonymity. Individual ages were not disclosed due to local regulations, preventing an accurate calculation of age means: however, expected average age in that class based on national curriculum data was 9 years. In this set of experiments, we used a sophisticated story variant described in section 6 that emphasizes both the special skill of Tuk (carving, which will play a critical role in the story ending), as well as the importance of family ties as the background for the narrative.

The experiment was run across two computer labs, one lab presenting the baseline version of the Tuk story, and the other the automatically generated variant with additional cues intended to aid each child's understanding of the various inferences involved. Three different questionnaires were used during these experiments: two during the animation itself, which was paused at some fixed stages to administer the questionnaire, and one post-hoc. They were based on both free text and multiple-choice questions testing children's understanding and inferences. The final questionnaire assessed the understanding of the story resolution, as well as how interesting and engaging the virtual narrative was rated by children.

Initially, we looked at the correctness of the objective multiple-choice questions. There was no statistically significant difference in the number of correctly answered multiple-choice questions between the control condition (M = 6.59, SD = 1.37) and the experimental condition (M = 6.69, SD = 1.67), $t(51)$ = -0.24, p = .81, *ns*).

However, one mid-story question ("What special skill does Tuk have?") resulted in very different responses between the control and experimental conditions, with only 11% correct answers in the baseline version. By contrast, 35% of children provided a correct answer after watching the alternate story, which constitutes a significant effect, $\chi^2(1) = 4.18$, p = .04 (two-tailed), Cramer's V = .28. The odds of correct response were thus 4.23 times higher in the alternate condition than in the baseline condition.

This observation is consistent with the timing of the cue presentation, as in the story variant the carving skills are presented shortly before the second questionnaire, as well as being emphasised (B). However, this goes beyond a simple recency effect, as the cue takes its significance in context: in the baseline story, an early presentation of the carving skills before some of the mid-story dramatic events took place may have rendered this presentation more anecdotal and less relevant.

This hypothesis is consistent with the other finding from our study, which is the significant association between experimental condition (baseline/variant) and whether a participant identified correctly Tuk's main goal as "becoming a hunter" in the final (post-hoc) questionnaire, $\chi 2(1) = 4.18$, p = .04 (two-tailed), Cramer's V = .28 (medium).

In the story variant, the presentation of the carving skills closer to their use in the narrative resolution is likely to have facilitated the inference. It also confers a greater agency to the Tuk character which may have emphasized his role in defeating the bear, hence facilitating the recall of Tuk's aspiration to become a hunter.

In this set of experiments we also questioned the acceptability

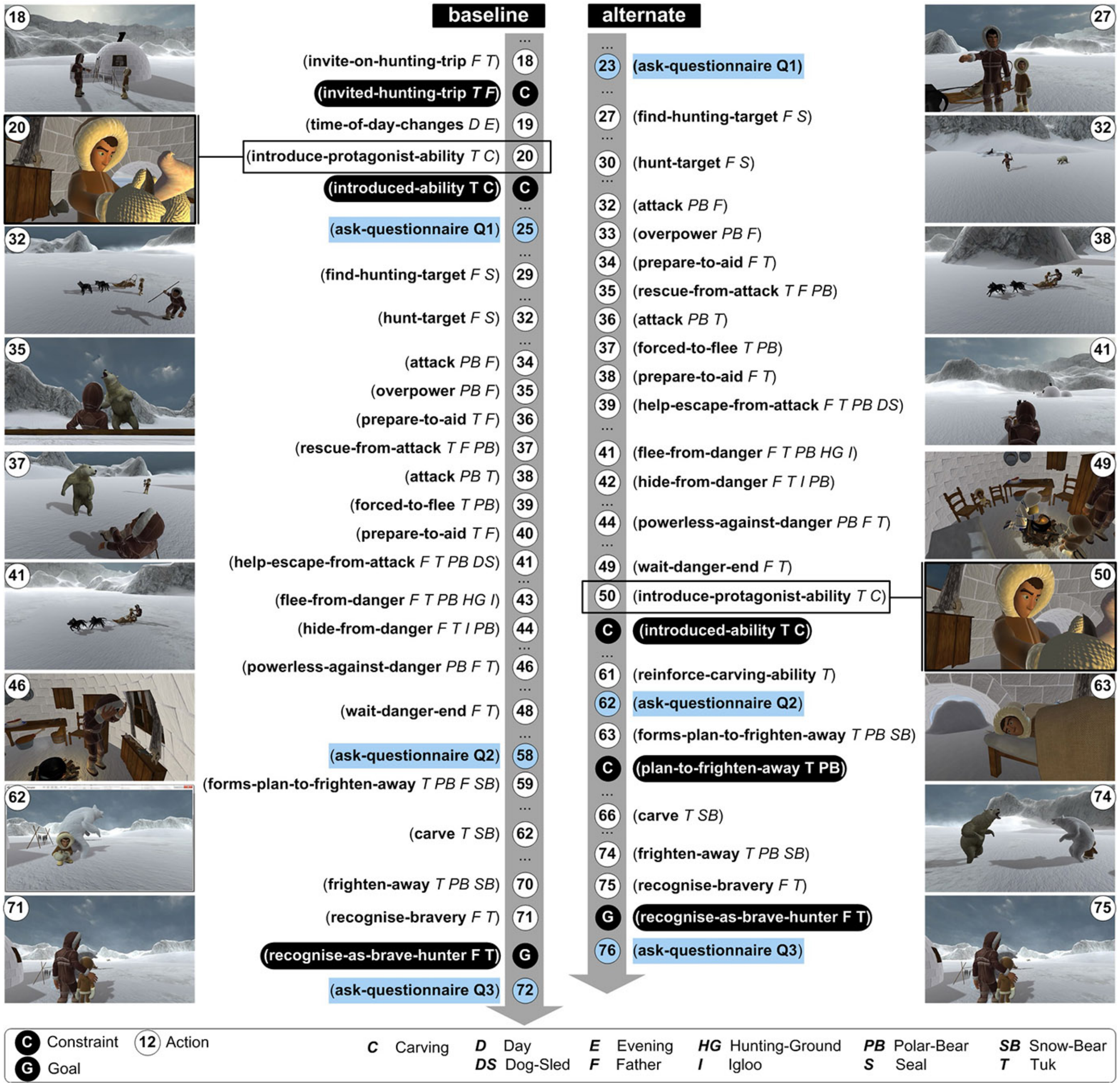


Figure 8: Example Narrative for the story of Tuk showing baseline (lhs) and an alternate variant (rhs). Narrative actions are numbered to show their relative positioning in the story, with the dedicated ask-questionnaire actions highlighted (blue). Constraints used to structure the narrative content are also highlighted (black). It can be seen that the cues related to Tuk's carving ability are included much earlier in the baseline narrative (action #20) in comparison to its later presentation in the alternate narrative (action #50). In addition in the alternate narrative the carving ability is also reinforced soon after (action #61). See text for further detail.

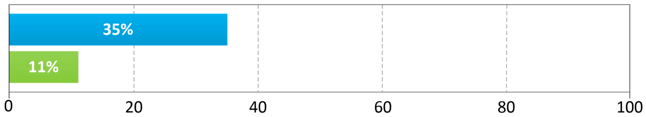


Figure 9: Experiment: participant responses to the mid-story question *"What special skill does Tuk have?"*. Only 11% of children who viewed the baseline version answered correctly, whereas 35% provided a correct answer after watching the alternate story with the carving skills cue strategically positioned. A significant effect.

of the virtual narrative to children, including several Likert-scale (1-5) questions assessing the story interest and its entertaining nature, such questionnaires being administered after the story was completed. These confirmed that children were able to successfully engage with the story, with the interest of the story being rated at a rather high average of 4.17 (*SD*=1.09) out of 5. There was no statistical difference for interest between the control story and the experimental condition (baseline 4.30 (*SD*=1.03), alternate 4.04 (*SD*=1.15), $t(51)=0.86$, $p=.39$, *ns*. Children were neutral on whether the story was "confusing" (*mean* 3.02, *SD*=1.47) and, importantly, there was no statistical difference between the control and experimental condition (baseline 2.74 (*SD*=1.46), alternate 3.31 (*SD*=1.46), $t(51)=-1.41$, $p=.16$, *ns*), suggesting that the added cues or modified scenes did not introduce confusion. It should be noted here that this corresponds to the subjective impression of children and does not test fine-grained understanding, which is the object of the above questionnaires.

7.2 Discussion

Our main objective for this first set of experiments was to assess system performance and acceptance rather than to contribute actual psychology findings with our first prototype. We still ensured that the experimental conditions were as close as possible to those for which the system has been developed in order to demonstrate the system potential. The fact that our early experiments yielded statistically significant results with non-trivial narrative configurations constitutes a first validation of the approach.

The sophistication of the story variants, which balance introduction of cues and narrative consistency, fully justifies the recourse to narrative generation as such results could hardly be obtained with the manual editing of animations, notwithstanding the ability of the system to generate a large number of different experimental conditions for a given set of cues. The example, illustrated in Fig. 8, is not able to fully illustrate this generative power due to lack of space, but is a direct property of narrative generation [29]. The alternate version of the Tuk story was similar in length to the original version (15 minutes) and, as noted above, the alternate story was shown to be as equally engaging as the original one, and was not perceived to make the story any more confusing or introducing additional comprehension difficulties. It should also be noted that, at 15 minutes, the total duration of the story is compatible with the state-of-the-art of narrative generation, which reinforces the ecological nature of the experiment.

The experiments clearly confirmed the ability of narrative generation to incorporate cues, which affect the understanding of the story without altering the narrative experience measured through engagement. In addition, the introduction of cues did not simply result in local inference facilitation but also had an impact on global aspects of story understanding. Furthermore, other effects were observed on the recognition of Tuk's main goal, suggesting that a complex set of cues could be introduced simultaneously.

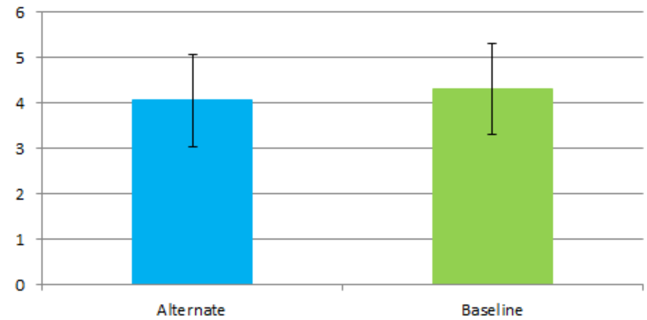


Figure 10: Narrative variants preserve story interest and engagement. Children participating in the experiment were asked to say how entertaining they found the narrative (rating on a 5 point Likert scale): engagement was high in both versions, with the baseline having a mean of 4.30 (*SD*=1.03), and the alternate 4.04 (*SD*=1.15), $t(51)=0.86$, $p=.39$, difference not significant.

8. CONCLUSIONS

We have shown that the narrative generation mechanism we introduced in this work has the ability to generate sophisticated variants of a given story, that emphasize specific presentation elements, or the ordering and distance between key events, while preserving the identity and consistency of a baseline plot. The ability to control variant generation is provided via an intuitive GUI. Hence the advantage of the narrative generation mechanism in this context over e.g. manual editing of story content and animations is clearly its ability to generate complex distributions of a limited number of cues or even to insert multiple cues, still preserving plot consistency. This generation also allows an experimenter to explore multiple variants prior to testing them with children, filtering them out on aesthetic or other criteria not covered by the generation mechanism itself.

Our demonstrator system was able to generate complete experimental variants realistic enough to stage understanding tests with the target children population. Results of the experiments have shown very good levels of acceptance and interest in the virtual narratives by children as well as some preliminary experimental results on the potential of narrative cues to study narrative inferences.

Overall we have shown that Interactive Narratives offer the potential to study narrative understanding independently of linguistic phenomena and linguistic abilities: lexical knowledge, anaphora, referential continuity to possibly investigate the existence of purely narrative phenomena. In addition another application would be to support narrative understanding experiments in children younger than the age of 6, with little or no reading abilities [10].

In future work we aim to research the ideal balance between 3D animations and language in terms of comprehension phenomena, by exploring various multimedia strategies, from spoken dialogues or captions to the insertion of silent-movie type panels introducing the main concepts during the story itself.

REFERENCES

[1] R. Aylett, J. Dias, and A. Paiva. An Affectively Driven Planner for Synthetic Characters. In *Proc. of 16th Int. Conf. on Automated Planning and Scheduling (ICAPS)*, 2006.

[2] R. Aylett, L. Hall, S. Tazzyman, B. Endrass, E. André, C. Ritter, A. Nazir, A. Paiva, G. Höfstede, and A. Kappas. Werewolves, cheats, and cultural sensitivity. In *Proc. of the*

13th Int. Conf. on Autonomous agents and multi-agent systems (AAMAS)*. IFAAMAS, 2014.

[3] D. B. Christian and R. M. Young. Comparing Cognitive and Computational Models of Narrative Structure. In *Proc. of the 19th National Conference on Artificial Intelligence (AAAI)*, 2004.

[4] A. Gerevini and D. Long. Plan Constraints and Preferences in PDDL3. Technical report, Department of Electronics for Automation, University of Brescia, Italy, 2005. http://www.cs.yale.edu/homes/dvm/papers/pddl-ipc5.pdf.

[5] A. C. Graesser, K. L. Lang, and R. M. Roberts. Question answering in the context of stories. *Journal of Experimental Psychology: General*, 120(3):254, 1991.

[6] A. C. Graesser, K. K. Millis, and R. A. Zwaan. Discourse comprehension. *Annual review of psychology*, 48(1):163–189, 1997.

[7] J. Hoffmann and B. Nebel. The FF Planning System: Fast Plan Generation throughHeuristic Search. *Journal of AI Research*, 14:253–302, 2001.

[8] A. Jhala and R. M. Young. Cinematic visual discourse: Representation, generation, and evaluation. *Computational Intelligence and AI in Games, IEEE Transactions on*, 2(2):69–81, 2010.

[9] P. Kendeou, C. Bohn-Gettler, M. J. White, and P. Van Den Broek. Children's inference generation across different media. *Journal of Research in Reading*, 31(3):259–272, 2008.

[10] P. Kendeou, P. van den Broek, M. J. White, and J. Lynch. Comprehension in preschool and early elementary children: Skill development and strategy interventions. *Reading comprehension strategies: Theories, interventions, and technologies*, pages 27–45, 2007.

[11] W. Kintsch. Meaning in context. *Handbook of latent semantic analysis*, pages 89–105, 2007.

[12] S. W. Mcquiggan, J. P. Rowe, S. Lee, and J. C. Lester. Story-based learning: The impact of narrative on learning experiences and outcomes. In *Intelligent tutoring systems*, pages 530–539. Springer, 2008.

[13] R. Melzack. *The Day Tuk Became a Hunter*. Macmillian, 1968.

[14] B. Mott, S. Lee, and J. Lester. Probabilistic goal recognition in interactive narrative environments. In *Proc. of the 21st Nat. Conf. on Artificial Intelligence*. AAAI Press, 2006.

[15] B. Mott, S. W. McQuiggan, S. Lee, S. Y. Lee, and J. C. Lester. Narrative-centered environments for guided exploratory learning. In *Proc. of the AAMAS 2006 Workshop on Agent-Based Systems for Human Learning*, 2006.

[16] B. W. Mott and J. C. Lester. Narrative-centered tutorial planning for inquiry-based learning environments. In *Intelligent Tutoring Systems*, pages 675–684. Springer, 2006.

[17] J. Porteous, M. Cavazza, and F. Charles. Applying Planning to Interactive Storytelling: Narrative Control using State Constraints. *ACM Transactions on Intelligent Systems and Technology (ACM TIST)*, 1(2):1–21, 2010.

[18] J. Porteous, M. Cavazza, and F. Charles. Narrative Generation through Characters' Point of View. In *Proc. of 9th Int. Conf. on Autonomous Agents and MultiAgent Systems (AAMAS 2010)*, 2010.

[19] J. Porteous, F. Charles, and M. Cavazza. NetworkING: using Character Relationships for Interactive Narrative Generation. In *Proc. of 12th Int. Conf. on Autonomous agents and multi-agent systems (AAMAS)*. IFAAMAS, 2013.

[20] M. O. Riedl and R. M. Young. Narrative Planning: Balancing Plot and Character. *Journal of AI Research*, 39:217–267, 2010.

[21] M. Si, S. C. Marsella, and D. V. Pynadath. THESPIAN: An Architecture for Interactive Pedagogical Drama. In *AIED*, pages 595–602, 2005.

[22] M. Si, S. C. Marsella, and D. V. Pynadath. Evaluating directorial control in a character-centric interactive narrative framework. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 1-Volume 1*, pages 1289–1296. International Foundation for Autonomous Agents and Multiagent Systems, 2010.

[23] N. L. Stein and C. G. Glenn. *New directions in discourse processing*, volume 2, chapter An analysis of story comprehension in elementary school children, pages 53–119. Norwood, NJ: Ablex, 1979.

[24] W. Swartout, R. W. Hill, J. Gratch, W. L. Johnson, C. Kyriakakis, C. Labore, R. Lindheim, S. C. Marsella, D. Miraglia, B. Moore, J. Morie, J. Rickel, M. Thiebaux, L. Tuch, R. Whitney, and J. Douglas. Toward the Holodeck: Integrating Graphics, Sound, Character and Story. In *Proc. of the 5th International Conference on Autonomous Agents*, Montreal, Canada, June 2001.

[25] Z. K. Takacs, E. K. Swart, and A. G. Bus. Benefits and pitfalls of multimedia and interactive features in technology-enhanced storybooks a meta-analysis. *Review of educational research*, 85(4):698–739, 2015.

[26] M. Vala, P. Sequeira, A. Paiva, and R. Aylett. Fearnot! demo: a virtual environment with synthetic characters to help bullying. In *Proc. of the 6th Int. Joint Conf. on Autonomous agents and multiagent systems (AAMAS)*. IFAAMAS, 2007.

[27] P. Van den Broek. Using texts in science education: Cognitive processes and knowledge representation. *Science*, 328(5977):453–456, 2010.

[28] P. Wolff and M. Zettergren. A vector model of causal meaning. In *Proc. of the 24th Annual Conf. of the Cognitive Science Society*, 2002.

[29] R. M. Young. Creating Interactive Narrative Structures: The Potential for AI Approaches. In *AAAI Spring Symposium in Artificial Intelligence and Entertainment*. AAAI Press, 2000.

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