Asking Hypothetical Questions About Stories Using QUEST

Rachelyn Farrell(✉), Scott Robertson, and Stephen G. Ware

Narrative Intelligence Lab, University of New Orleans, New Orleans, LA 70148, USA
{rfarrell,sgware}@uno.edu, srobert6@my.uno.edu

Abstract. Many computational models of narrative include representations of possible worlds—events that never actually occur in the story but that are planned or perceived by the story's characters. Psychological tools such as QUEST are often used to validate computational models of narrative, but they only represent events which are explicitly narrated in the story. In this paper, we demonstrate that audiences can and do reason about other possible worlds when experiencing a narrative, and that the QKSs for each possible world can be treated as a single data structure. Participants read a short text story and were asked hypothetical questions that prompted them to consider alternative endings. When asked about events that needed to change as a result of the hypothetical, they produced answers that were consistent with answers generated by QUEST from a *different* version of the story. When asked about unrelated events, their answers matched those generated by QUEST from the version of the story they read.

Keywords: Hypothetical reasoning · Planning · Possible worlds · QUEST

1 Introduction

Narrative theorists often analyze stories in terms of *possible worlds* [1–3]. A story's meaning may arise not only from the events which did happen but also from the events which did not happen or might have happened instead.

This theory of possible worlds is informed by the idea that when a speaker communicates a narrative, the hearer is an active participant [4]. The audience engages in a complex cognitive process of constructing a mental model of the story while consuming it. Cognitive scientists have demonstrated that the hearer constantly updates this mental model by revising working memory [5], shifting the focus of attention [6], making assumptions about missing information [7], and making inferences about the future [8], among other tasks.

The possible worlds theory has recently informed computational models of narrative as well. Riedl and Young [9] developed a formal, generative, plan-based representation of stories which ensures that every action in the plan can be explained not only in terms of the author's goals for the story as a whole,

F. Nack and A.S. Gordon (Eds.): ICIDS 2016, LNCS 10045, pp. 136–146, 2016.
DOI: 10.1007/978-3-319-48279-8_12

but also in terms of the individual goals of the characters who are acting. This model made the story's characters more believable, but it was unable to represent characters' plans which failed or were only partially completed, which is an essential element of conflict [10].

Ware and Young [11] solved this problem by extending that representation to model other possible worlds. The Glaive narrative planning algorithm [12] treats the search space of a planning problem as a Kripke structure [1], where each branch represents a different way the story might have unfolded. Each event in the story still needs to be explained in terms of character goals, but it is sufficient for that explanation to appear in a different possible world from the actual story. They give an example from *Indiana Jones and the Raiders of the Lost Ark* (1981), in which Jones excavates the Ark of the Covenant only for it to be stolen by the antagonists. We can explain that Jones excavated the Ark with the intention of taking it to safety, even though this plan fails and we never see it happen. It is enough to know that there exists a possible world in which Jones took the Ark to safety. Similar possible worlds reasoning has been proposed for representing other narrative phenomena, such as the differing beliefs of characters [13].

These computational models of narrative are often evaluated using psychological tools such as QUEST [14], which represents the audience's cognitive state in terms of question-answering ability. However, these tools only reason about those events which are actually narrated, making it impossible to apply them to models which require the reader to reason about other possible worlds. Reasoning about possible alternatives is also an essential task for analyzing interactive narratives.

In this paper, we demonstrate that QUEST can be used to model how audiences reason about other possible worlds when experiencing a narrative. After reading a short text story, participants were asked hypothetical questions that made some events in the story impossible or enabled new events which were previously impossible. By translating these hypothetical prompts into modifications on the QUEST knowledge structure, we generated new structures that represented the hypothetical worlds that the audience was prompted to imagine. When asked hypothetical questions, participants gave answers that corresponded to the QUEST model of the *alternate* version of the story. We consider this investigation to be preliminary evidence that empirical tools such as QUEST can be applied to validate computational models which are based on possible worlds reasoning.

2 Related Work

QUEST is a framework developed by cognitive scientists to predict how adults answer open-ended questions about finite sets of information. Short text narratives can be encoded as a QUEST Knowledge Structure (QKS), a directed graph composed of nodes whose content is a simple sentence statement about some element of the story. QUEST defines five types of nodes and six types of edges that correspond to different types of content and relationships in a narrative. QUEST also defines graph search procedures for answering *why? how? when? what enabled?* and *what are the consequences of?* questions. Graesser et al. [14] demonstrated that when human subjects are given a question and asked to rank a set of answers, these search procedures reliably predict Goodness of Answer (GoA) and can serve as a proxy for elements of human narrative understanding.

QUEST has been used to validate computational models of narrative. Christian and Young [15] developed a mapping of plan-based models of narrative onto QKSs. This mapping was later updated by Riedl and Young [9] to incorporate character goals. Recently Cardona-Rivera et al. [16] introduced a variant mapping to improve the accuracy of *why?* questions. These mappings facilitated the validation of several plan-based computational models of narrative. For example, Riedl and Young [9] used it to demonstrate that intentional planning generated more believable stories by showing that human subjects consistently choose answers with higher goodness (according to QUEST GoA measures) when reading stories generated by their planner.

The trouble with QUEST is that a QKS is defined only for those events which are actually narrated. Events which did not happen or might have happened instead are not represented, so one cannot ask questions about them or consider them when answering other questions. In short, QUEST lacks support for hypothetical questions, and thus is of limited usefulness for validating models based on possible worlds.

Boyd [17] argues that storytelling contributed to human evolutionary success because (among other factors) it exercised the ability to reason hypothetically. Bruner [2,18] claims that narrative is the primary way that we structure experiences into an understandable reality and that our ability to perceive the actual world is based on a general ability to imagine possible worlds. Hypothetical reasoning is essential to narrative, and it has prompted numerous theorists to develop systems of analysis based on the idea that when someone experiences a narrative he or she can reason about a network of possible worlds [13,19–21]. These theories are based on Kripke's semantics for modal logic [1] about what is necessary, possible, and impossible.

Some work has been done related to cognitive tools and hypothetical reasoning. Graesser and Olde [22] studied how people ask questions (including *what if?* and *what if not?* questions) when learning how a device works and how it can malfunction. Gerrig and Bernardo [8] demonstrated that when reading short James Bond stories the audience experienced higher suspense when the number of foreseeable possible worlds that are good for the protagonist decreases. This model of suspense was leveraged by Cheong and Young [23] in the *Suspenser* story generation system.

We propose that we can incorporate events from other possible worlds into a QKS by treating multiple QKSs as if they were one structure, and thereby extend QUEST to be able to reason about events that never actually happened. We do this by translating hypothetical prompts into insertions or deletions of

QUEST nodes on a QKS. For example, to represent the possible world prompted by the hypothetical "What if X had happened?", we add a node X to the QKS. Similarly, for "What if X had not happened?" we remove node X. We demonstrate that these models of possible worlds can accurately predict the audience's answers to hypothetical questions involving events that were not explicitly narrated.

3 QKS Mapping

Although we mapped each story to a QKS using Riedl and Young's complete mapping algorithm [9], we only ask one type of question in our experiment—*how?*—and therefore limit our discussion to the node and arc types accessed by the arc-search procedure for *how?* questions. The mapping introduced by Cardona-Rivera et al. [16] is identical to Riedl and Young's for these kinds of questions.

Figure 1 gives an example of a Goal-node hierarchy, in which subordinate goals are connected to super-ordinate goals via Reason arcs. This example shows how the character Sarah plans to achieve her goal of being a parent. The Reason arcs between the Goal nodes indicate that her goal *Sam and Sarah get married* is explained by her higher-level goal *Sam and Sarah have a baby named Hope*, which is in turn explained by her top-level goal, *Sarah is a parent*.

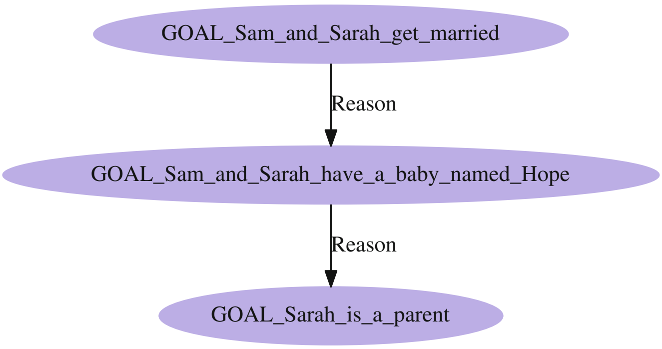

Fig. 1. Example QKS goal node hierarchy

QUEST answers questions by first identifying the type of question being asked and the node being queried, and then performing a breadth-first search of the QKS starting at the queried node and traversing only the arcs that are legal for that question type. The set of candidate answers to the question is the set of all nodes reached by this search. The arc-search procedure for the *how?* question allows backward Reason and Manner arcs, so if we asked the question, "How does Sarah become a parent?", we would generate the answer set {*Sam and Sarah have a baby named Hope*, *Sam and Sarah get married*}.

Our goal node hierarchies were created by the following subset of the mapping algorithm:

– For every action taken in the story, we create a Goal node for each character who must consent to take that action. For example, the event *Sam and Sarah get married* has two corresponding Goal nodes—one representing Sam's goal for them to get married, and one for Sarah's.
– A *causal link* is a structure in a plan that represents a dependency between two actions; the first action establishes some necessary precondition for the second action. For each causal link in the plan, we add a Reason arc from the Goal node for the first action to the Goal node for the second action.

4 Experimental Design

We tested the ability of an audience to reason hypothetically by having them read a story and consider hypothetical situations. We hypothesized that when those situations made important events possible or impossible, they would answer as if they had read an alternate version of the story where that hypothetical was the reality, demonstrating that they can reason about other possible worlds even when the author does not explicitly narrate them. We also sought to observe the absence of this phenomenon. When asked to consider a hypothetical that would not affect the answer to a question, they would answer consistently with the story they read.

4.1 Materials

We used the Glaive narrative planning system [12] to generate four stories. Each had the same beginning but a different ending, representing four possible worlds. Each of these stories was translated into a QKS. QUEST search procedures were used to generate candidate answers to the questions we intended to ask. Human subjects were shown one of the four stories and asked to consider a hypothetical scenario. Subjects then answered questions by ranking a set of possible answers from best to worst. This set of answers was generated by combining QUEST's candidate answers to the same question across all four stories (the story that subject read, plus the other stories they did not read). The order in which readers ranked answers revealed which of the four stories they were reasoning about.

Figure 2 gives the text of the four stories. In all versions, Sarah has two important character goals which can be satisfied in two different ways: she wants to get a job and she wants to be a parent. In all versions, she applies to work at Google, indicating that she plans to achieve her goal of having a job by working for Google. In stories A1 and B1, Google offers her a job and she accepts, but in versions A2 and B2, Google never offers her the job. Instead, she applies for and accepts a different job at Home Depot. In stories A1 and A2, Sarah achieves her goal of becoming a parent by having a child with her spouse. In stories B1 and B2, she achieves this goal by adopting a child.

William visits an adoption agency. William adopts a child named Jude. William and Candace fall in love. William and Candace get married. William and Candace have a baby named Sarah. A boy named Sam grows up. Sarah grows up. Sam applies for a job at Home Depot. Home Depot offers Sam a job. Sam takes the job at Home Depot. Sam and Sarah fall in love. Sam and Sarah get married. Sarah applies for a job at Google.	**All Stories**
Google offers Sarah a job. Sarah takes the job at Google. Sam and Sarah have a baby named Hope.	**Story A1**
Google offers Sarah a job. Sarah takes the job at Google. Sarah visits an adoption agency. Sarah adopts a child named Abe.	**Story B1**
Sarah applies for a job at Home Depot. Home Depot offers Sarah a job. Sarah takes the job at Home Depot. Sam and Sarah have a baby named Hope.	**Story A2**
Sarah applies for a job at Home Depot. Home Depot offers Sarah a job. Sarah takes the job at Home Depot. Sarah visits an adoption agency. Sarah adopts a child named Abe.	**Story B2**

Fig. 2. Story text for all four variations

When participants read a story in which Sarah got the job at Google (A1 and B1), they were asked to consider the hypothetical, "What if Google had not offered Sarah a job?" If they read a story where she did not get the job (A2 and B2), they were asked to consider, "What if Google *had* offered Sarah a job?"

4.2 Questions

To summarize the experiment so far, participants read one of four variants of the same story and were asked to consider a hypothetical. They then answered two questions, one whose answer should be affected by the hypothetical and one whose answer should not be affected.

The first question was, "How would Sarah have achieved her goal of having a job?" The answers we provided included all candidate answers generated by QUEST using the *how* arc search procedure on all story versions:

- Sarah takes the job at Google.
- Sarah applies for a job at Google.
- Sarah takes the job at Home Depot.
- Sarah applies for a job at Home Depot.

We expected subjects to prioritize the answers that came from a version of the story they *did not* read. If they read a story where Sarah was offered the job at Google, and were asked to consider what would have happened if she had not gotten that offer, we expected them to say she would seek a job elsewhere. If they read a story where she did not get the offer, and were asked to consider what would have happened if she *had* gotten the offer, we expected them to say she would take the job at Google instead.

The second question was, "How would Sarah have achieved her goal of becoming a parent?" The candidate answers were:

- Sam and Sarah have a baby named Hope.
- Sam and Sarah get married.
- Sarah adopts a child named Abe.
- Sarah visits an adoption agency.

Here we expected participants to prioritize the answers that came from the version of the story they *did* read. The hypothetical affects her career goal by making her ideal career possible or impossible; however it does not directly affect her family goal. If subjects read a story where she had a child with her spouse, we expected them to prioritize that answer despite the hypothetical. Likewise, if they read a story where she chose to adopt a child, we expected them to prioritize that answer, despite the hypothetical. This second question was important because we not only want to demonstrate that people can change their mental model of the story when they have to, but also that they do not change it when they don't have to.

4.3 Controlling for Story Content and Response Quality

One potential complication that arises from hypothetical reasoning is that there is a potentially infinite number of other ways something might happen. We attempted to control for this in two ways. Firstly, we do not allow open-ended responses but rather ask users to rank a pre-generated set of candidate responses. Secondly, the beginning common to all stories was designed to introduce all the people, places, things, and types of actions that could occur. The four possible endings all reuse the existing entities and actions. For example, the beginning of the story includes examples of a child being born and a child being adopted. Thus, even if subjects read a story where Sarah's child was born, they knew that adoption was also possible.

Subjects were recruited via Amazon Mechanical Turk. To account for the large amount of noise, we asked all participants to complete an initial training task with a different story (without being asked to consider any hypothetical situations). This training task may have had a priming effect that influenced in what order subjects ranked their answers from the same story. This was an acceptable risk because this experiment is not designed to validate QUEST search procedures or question answering tasks. We consider that task already accomplished by Graesser et al. [14]. We are only interested in testing whether or not a hypothetical situation causes subjects to report answers from the same story or a different story.

Finally, as a further quality control, each subject was asked two non-hypothetical comprehension questions after they read the story. For each question, they were instructed to rank a given set of answers from best to worst. Of the answers we provided, two were produced by QUEST's arc-search procedure for the given question, and the other two were chosen arbitrarily from the other events in the story, such that they had no relation to the question being asked. We assume that if the reader fully comprehended the story they would rank the two answers produced by QUEST higher than the two unrelated events. Again, this assumption arises from previous work on QUEST and is not the goal of this study. If they did not rank the answers for both questions accordingly, we assumed they did not pay close attention to the story and discarded their data as noise.

5 Results

We recruited 180 participants in total and retained 88 responses after discarding noisy data. Table 1 shows the number of responses from each story version and the distribution of responses to each question. For example, of the 26 participants who read version A1, there were 5 who answered the *job* question using events from the version-1 stories (A1/B1), and 21 who answered it with events from the version-2 stories. Similarly, there were 23 who answered the *child* question using events from version A, and 3 who answered it with events from version B.

Table 1. Response distribution for each question and all story versions

		Job question		Child question	
Story Read	Number of responses	Answered 1	Answered 2	Answered A	Answered B
A1	26	5	21	23	3
B1	22	5	17	10	12
A2	20	20	0	18	2
B2	20	18	2	8	12

Hypothesis #1. When presented with a hypothetical that makes critical events become impossible, people will answer using events from a story other than the one they read.

In other words, for participants who read story A1 or B1, we expect the highest ranked answer to the hypothetical *job* question to come from story A2 or B2. The binomial exact test supports this hypothesis with $p < 3.085e^{-5}$.

Hypothesis #2. When presented with a hypothetical that makes new events possible that were not possible before, but that were part of a character's plan, people will also answer using events from a story other than the one they read.

In other words, for participants who read story A2 or B2, we expect the highest ranked answer to the hypothetical *job* question to come from story A1 or B1. The binomial exact test supports this hypothesis with $p < 7.467e^{-10}$.

The first two hypotheses consider the two cases individually and test how the results deviate from the null hypothesis that there is an underlying binomial distribution. However, there are many reasons outside of our experiment why a binomial distribution may not hold in these cases. To provide further evidence that we are observing a real effect, we now group hypothesis #1 and #2 together into a more general hypothesis.

Hypothesis #3. When presented with a hypothetical about a critical event, people will answer using events from a story other than the one they read.

Fisher's exact test shows that there is a significant association ($p < 2.823e^{-13}$) between the story that participants read and the story that their answers came from; that they tended to select answers from the opposite story. There are several ways to measure effect size when using Fisher's exact test. The odds ratio for this contingency table is ≈66.96, meaning there are about 67 to 1 odds that users chose answers from the alternate story.

Hypothesis #4. When presented with a hypothetical about a *non-critical* event, people will answer using events from the story they read.

In other words, we expect all participants to answer the *parent* question using answers from the story they read, since Sarah's plan for becoming a parent is unrelated to the event of Google offering her a job.

Fisher's exact test supports this with $p < 3.497e^{-6}$, with an odds ratio of $\approx$10.6. We conclude that readers are indeed only using answers from a different story when it makes sense to.

6 Conclusion and Future Work

Computational cognitive tools like QUEST can be a valuable resource for validating formal models of narrative, especially plan-based models which have a fairly direct mapping from plan to QKS. However, QUEST only represents the actual world. Many phenomena, especially in interactive narratives, require reasoning about alternative possible worlds.

In this paper we have provided preliminary evidence that, for certain kinds of content, QUEST can be extended to reason about other possible worlds by treating multiple QKSs as if they were one graph. Given the appropriate prompts, users realize when events have become impossible and can reason about alternative ways to complete the story. Likewise, when new events are introduced, they can update their expectations to incorporate these new events. They are also able to distinguish when these hypotheticals are relevant and avoid updating their expectations when it is unnecessary.

It is important to point out that we have only used a single story domain and only tested *how?* questions, which use a limited subset of QUEST's available features. We chose to focus on *how?* questions because reasoning about the ways characters achieve their goals is frequently needed in plan-based interactive stories. Obviously the value of our results can be enhanced through replication in different domains. It will also be important to study whether or not other kinds of questions, like the *why?* questions used by Riedl and Young [9] and Cardona-Rivera et al. [16], are affected by hypothetical reasoning.

We suspect that the audience's mental model contains information about many possible worlds, not just the actual world that corresponds to the story as narrated. By incorporating other possible worlds into a QKS, it more closely reflects the reader's mental model and can reliably answer certain kinds of *what if?* and *what if not?* questions. Narrative generation systems which reason about these other possible worlds can better reflect the human comprehension process. We hope these results will be valuable to other researchers using empirical tools such as QUEST to study computational models of narrative.

References

1. Kripke, S.A.: Semantical considerations on modal logic. Math. Logic Q. **9**, 67–96 (1963)
2. Bruner, J.S.: Actual Minds, Possible Worlds. Harvard University Press, Cambridge (1986)
3. Ryan, M.L.: Possible worlds (2016)
4. Gerrig, R.J.: Experiencing Narrative Worlds: On the Psychological Activities of Reading. Yale University Press, New Haven (1993)

5. Baddeley, A.D., Hitch, G.: Working memory. Psychol. Learn. Motiv. **8**, 47–89 (1974)
6. Zwaan, R.A., Radvansky, G.A.: Situation models in language comprehension and memory. Psychol. Bull. **123**, 162 (1998)
7. van den Broek, P.: Causal inferences and the comprehension of narrative texts. Psychol. Learn. Motiv. **25**, 175–196 (1990)
8. Gerrig, R.J., Bernardo, A.B.I.: Readers as problem-solvers in the experience of suspense. Poetics **22**, 459–472 (1994)
9. Riedl, M.O., Young, R.M.: Narrative planning: balancing plot and character. J. Artif. Intell. Res. **39**, 217–268 (2010)
10. Herman, D.: Story Logic. University of Nebraska Press, Lincoln (2002)
11. Young, R.M., Harrison, B., Roberts, D.L.: A computational model of narrative conflict at the fabula level. IEEE Trans. Comput. Intell. Artif. Intell. Games **6**, 271–288 (2014)
12. Young, R.M.: Glaive: a state-space narrative planner supporting intentionality and conflict. In: Proceedings of the 10th AAAI International Conference on Artificial Intelligence and Interactive Digital Entertainment, pp. 80–86 (awarded Best Student Paper) (2014)
13. Ryan, M.L.: Possible Worlds, Artificial Intelligence, and Narrative Theory. Indiana University Press, Bloomington (1991)
14. Graesser, A.C., Gordon, S.E., Brainerd, L.E.: QUEST: a model of question answering. Comput. Math. Appl. **23**, 733–745 (1992)
15. Christian, D.B., Young, R.M.: Comparing cognitive and computational models of narrative structure. In: Proceedings of the 19th National Conference of the American Association for Artificial Intelligence, pp. 385–390 (2004)
16. Cardona-Rivera, R.E., Price, T., Winer, D., Young, R.M.: Question answering in the context of stories generated by computers. Adv. Cogn. Syst. **4**, 227–245 (2016)
17. Boyd, B.: On the Origin of Stories: Evolution, Cognition, and Fiction. Harvard University Press, Cambridge (2009)
18. Bruner, J.: The narrative construction of reality. Crit. Inq. **18**, 1–21 (1991)
19. Lewis, D.: Truth in fiction. Am. Philos. Q. **15**, 37–46 (1978)
20. Eco, U.: The Role of the Reader: Explorations in the Semiotics of Texts, vol. 318. Indiana University Press, Bloomington (1984)
21. Pavel, T.G.: Fictional Worlds. Harvard University Press, Cambridge (1986)
22. Graesser, A.C., Olde, B.A.: How does one know whether a person understands a device? the quality of the questions the person asks when the device breaks down. J. Educ. Psychol. **95**, 524 (2003)
23. Cheong, Y.G., Young, R.M.: Suspenser: a story generation system for suspense. IEEE Trans. Comput. Intell. Artif. Intell. Games **7**, 39–52 (2015)