Evaluating the Use of Action Templates to Create Suspense in Narrative Planning

Matthew William Fendt,   David Roberts,   R. Michael Young

Digital Games Research Center
Department of Computer Science
NC State University
Raleigh, NC 27695-8206

Abstract

The IRIS story generation system is a narrative planner that is designed to create suspenseful stories at the story-action level. It uses a set of action templates that are injected into the planning process to create suspense around a given goal. The action templates were evaluated independent of the rest of the IRIS system against a series of other techniques designed to create suspense. Results show that it is possible to use these set of action templates to create localized moments of suspense in a story at the story-action level without the need for discourse-level suspense.

Introduction

In recent years, the narrative aspect of games has grown in importance to game developers and players. Games like Fallout: New Vegas and Mass Effect 3 extol the importance of player decisions on the outcome of the story. This emphasis on narrative provides an opportunity to utilize the benefits of computer science by improving the game author's ability to incorporate his or her story into the game. Researchers seek to dynamically generate content, model player gaming behavior, influence player agency and decision making, provide a tailored game experience, and more, all in the interest of improving the narrative game experience for the player.

The IRIS story generation system (Fendt 2010), too, seeks to improve the player's narrative experience. One of the system's components is a small set of action templates that are introduced in the story plan creation process to create localized moments of story action-level suspense. Story-action level refers to the selection and combination of the story actions themselves, independent of how they are told or the medium in which they are presented. These action templates were evaluated independent of the rest of the IRIS system, in order to validate their effectiveness in creating suspense. The IRIS system will use these templates to produce a story outline which contains suspenseful action sequences. The output provides game designers an alternative to writing all of the suspenseful story moments manually. Additionally, the outline nature of the output of IRIS allows for flexibility in its instantiation into the game world, allowing choice of how the suspense is realized in game. An overview of the IRIS system is shown in Figure 1.
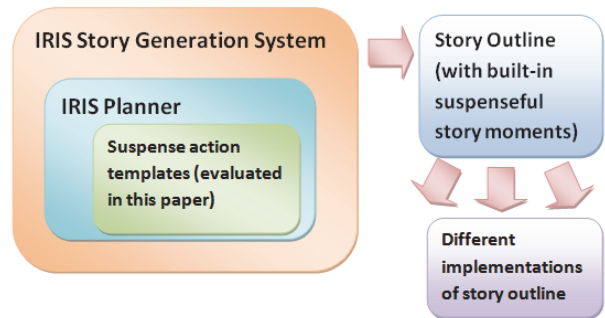
Figure 1: The IRIS story generation system. The part of the system that is being evaluated in this paper is the set of suspense action templates. They are used in the story planning process to produce plans whose execution will be suspenseful. The plans are used to create a story outline containing suspense that can then be instantiated in different media.

Over the course of two experiments, four different suspense creation processes were compared: the IRIS action templates, an operationalization of a case-based reasoning system with narrative properties called MINSTREL (Turner 1994), a human author with creative writing experience, and a logically consistent random baseline. Each process created a set of suspenseful story fragments given the same domain of a murder mystery and a grammar from which to produce the story fragments. Participants evaluated pair-wise comparisons of the different processes with regards to suspense. We found that stories produced by the human author and by most of the IRIS fragments, particularly those involving a disparity of knowledge between the audience and the protagonist, were significantly more suspenseful than the random story fragments. Additionally, the IRIS action templates did not perform significantly differently than the human author. This validates the IRIS action templates as a viable means of creating suspense at the story-action level.

Related Work

Suspense has been defined multiple different ways (Cheong and Young 2008), (Rodell 1943), but the definition that we will use is: "the feeling of excitement or anxiety that audience members feel when they are waiting for some-

thing to happen and are uncertain about a significant outcome" (Cheong and Young 2008). Suspense can be created many different ways. Some ways involve having the audience identify and empathize with the protagonist (Vorderer 1996), (Zillmann 1996). Other ways to build suspense come at the discourse, or narration, level, such as showing less detail and allowing the audience's imagination to fill in the detail (Wulff 1996) or having the audience anticipate failure, even if the possibility of failure is actually low (Wulff 1996). These ways of creating suspense are difficult to achieve at the story-action level, however, since most of them involve carefully crafted discourse. One way to create suspense at the story-action level is to demonstrate that there are only a few paths to success of the outcome, and then systematically cut off those paths (Cheong and Young 2008), (Gerrig and Bernardo 1994). For example, the hero is trying to get off an island with an erupting volcano. The hero tries to get to the helicopter, but finds that someone flew off in it. Then, the hero tries to cross the bridge, only to find that it has burnt down. The dwindling options produce a feeling of suspense in the reader because the success of the hero getting off of the island is uncertain. This feeling of suspense is partially influenced by the suspense intrinsic in the sequence of the story actions themselves. Discourse can also be used to create suspense; however, in this paper we focus on the former. This story-action level method for creating suspense is the backbone of the action templates evaluated in this paper.

There have been relatively few systems that introduce suspense into story generation. Besides IRIS, two story generation systems that have included suspense are Suspenser and MINSTREL. Suspenser (Cheong and Young 2008) is a planning system that generates suspenseful stories. It generates suspense at the discourse level. Suspenser first uses a planner to produce the story events. It then iteratively increases suspense by replacing and reordering story events. However, since Suspenser generates suspense at the discourse level, it may not be generating stories that are intrinsically suspenseful, but rather adding suspense to the stories after they have been generated. Suspenser also is not guaranteed to generate high suspense stories, since it can get stuck in a local maximum when adding suspense. The discourse-level suspense created by Suspenser is complimentary to the story-action level suspense our IRIS system creates. In future work, the two approaches could be combined; however, that is beyond the scope of this paper.

MINSTREL (Turner 1994) is a case-based reasoning system that creates stories that have, among other properties, suspense. It creates suspense both at the story-action level and the discourse level. MINSTREL has a rigid model that determines when and how to create this suspense. The only appropriate time to create suspense in its model is in a scene that the author wants to emphasize, and when the character's life is in danger. MINSTREL provides a framework for when, when not, and how to create suspense around an outcome. Its weakness is the simplistic way in which each of these criteria are determined. This is likely due to the fact that suspense is just one of several narrative phenomena that the system is trying to capture. Additionally, it only provides this one formula to generate suspense.

Description of Suspenseful Action Templates

IRIS (Fendt 2010) is a narrative planner that uses a Drama Manager and belief/desire/intention mental modeling with intention revision to produce suspenseful stories. This paper reports the results of two experiments on the suspense-producing components of the IRIS system and shows that a set of action templates is sufficient to create story-action level suspense in story fragment generation. IRIS produces local moments of suspense around certain goals at the story-action level. The long-term goal for the IRIS system is to take these templates and produce a story outline that is created using these templates to create suspense. The story outline can then be adapted to many different discourses in media such as text-based and three dimensional games.

The IRIS suspense templates were created to cover the different ways suspense can be enumerated within a specific definition of suspense. The templates were designed to create suspense at the story-action level. Within the story-action level, the IRIS system creates suspense from protagonist plan failure. The protagonist's plan will fail if one of the preconditions of the actions in its plan does not hold. There are three ways that the precondition can fail to hold.

The first way a precondition can fail to hold is an inconsistency between the protagonist's beliefs and the actual state of the world. This inconsistency can be unknown to the audience too, so that the audience is surprised by the plan failure, or known to the audience, resulting in dramatic irony.

The second way a precondition can fail to hold is if the system deliberately modifies the state of the world to introduce plan failure. In this case, a precondition of the protagonist's plan was true up to a point, but then was reversed by the system. This is analogous to initial state revision in planning (Riedl and Young 2005). This modification is unknown to the protagonist, but also unknown to the audience. It is kept secret from the audience because the change is a retroactive modification to the initial state of the world which seemingly has no perpetrator, so it would look unnatural for the audience if they saw it changed.

The third way a precondition can fail to hold is if an antagonistic agent takes action to thwart the precondition in the protagonist's plan. This agent could be another character or nature. The action may or may not be observed by the protagonist. It is observed by the audience, because otherwise this condition would be a subset of the first condition where, unknown to the protagonist and the audience, the state of the world is inconsistent with the protagonist's beliefs.

The high level Drama Manager actions from Figure 2 were translated into the suspense templates shown in Figure 3. This translation was made by enumerating the way these actions could fail using the three types of plan failure described above. The suspense templates that were not rated as suspenseful in the experiment are indicated by an asterisk.

Initial Experimental Design

An initial experiment evaluated the IRIS templates against multiple other suspense creation processes: three human authors, an operationalization of the MINSTREL system, and a logically and temporally consistent random baseline. The

Successful Action, SA(a): A character successfully performs an action a .

Failed Action, FA(a): A character attempts action a and fails.

Conflicting Action, CA(a): A character successfully performs an action a which will interfere with the protagonist's plan.

Hidden Alteration, HA(a): A change to the world state a is made without the characters' or audience's knowledge.

True Observation, TO(a): A character makes a correct observation about some state of the world a.

True Utterance, TU(a): One character truthfully informs another character about some state of the world a.

Discourse Statement, DS(a): A statement a about the world that is made to the audience but not the characters.

Figure 2: The high level actions in the IRIS system that can be combined to create suspense.

Cut Off Path
∗ 1) CA(a), TO(a)
∗ 2) FA(a), TO(a)
3) CA(a), withhold TO(a), FA(b), TO(b)
4) CA(a) (without character present), FA(b), TO(b)
Introduce New Info But Then Cut off Path
∗ 5) TU(a), HA(a), FA(b), TO(b)
∗ 6) TO(a), HA(a), FA(b), TO(b)
7) TU(a), CA(b), withhold TO(b), FA(c), TO(c)
8) TO(a), CA(b), withhold TO(b), FA(c), TO(c)
9) TU(a), CA(b), (without character present), FA(c), TO(c)
10) TO(a), CA(b) (without character present), FA(c), TO(c)
Dramatic Irony
∗ 11) DS(a), ... , plan failure due to world condition in DS(a)

Figure 3: Action templates that combine the high level actions from Figure 2 into suspenseful combinations of actions. The templates with an asterisk were not rated as suspenseful in the evaluation.

hypotheses were that the human authors and the IRIS fragments would be rated higher than the random baseline, and that the IRIS fragments would be rated comparably to the human authors. This rating would be done by participants reading the fragments and selecting which fragment, if any, was more suspenseful.

In the initial experiment, the story creation processes were given a bank of English sentences to use in story fragment creation. The domain was a knight trying to slay a dragon in a medieval setting. Example sentences included "The knight attacks the dragon," and "The wizard tells the knight where the dragon is hiding." Some simple rules were also given to ensure that the story would be logically and temporally consistent. This was done to ensure that that participants did not rate the fragment poorly because it did not make sense.

The study was conducted online in HTML with JavaScript to help handle the selection and display of the story fragments. Participants were recruited by snowball sampling through email, message boards, in person, and on Facebook.

The participants were first given a link to the consent form. Then, they were given a demographic survey asking their age and gender. In the demographic survey and the comparisons, the participants always had the option to leave any question unanswered. Next, they were given the background information about the domain of the story.

The participants were then given a series of sentence fragment pairs. They were asked to read each fragment and compare which was more suspenseful. They were given four options: Story Fragment A was suspenseful and Story Fragment B was not, Story Fragment B was suspenseful and Story Fragment A was not, Both are suspenseful, or Neither are suspenseful. After selecting their choice hitting the Next button, they were shown another pairing.

Four hypotheses were tested:

Hypothesis 1: The story fragments generated with the IRIS action templates will be comparably suspenseful to the fragments generated by the human authors selecting from the same set of story actions. To ensure that the story fragments that the action templates and the human authors generate are able to be compared, they both compose their fragments from the same set of story actions. This puts a limitation on the human author since he or she can only use a limited set of story actions instead of an unlimited vocabulary. If this hypothesis holds, then the action templates will have been shown to be sufficient to generate suspense at the story-action level comparable to a human author selecting from the same set of story actions.

Hypothesis 2: The story fragments generated with the IRIS action templates will be more suspenseful than the randomly-generated baseline. It needs to be shown that the action templates are an improvement over no method of creating suspense at all.

Hypothesis 3: The story fragments generated by the human authors will be more suspenseful than the randomly-generated baseline. This demonstrates that the human authors were actually able to create more suspense than random combinations of story events.

Hypothesis 4: The story fragments generated with the IRIS action templates will be at least as suspenseful as the MINSTREL formula. Since the IRIS action templates are more complicated to implement and apply than the MINSTREL formula, it is desirable that they perform at least as well as the simpler MINSTREL approach.

Initial Experimental Results

117 people completed the demographic survey and were presented with the experiment. There were 34 women with an average age of 33.6 +/- 15.2, and 83 men, with an average age of 26.2 +/- 9.8.

Table 1 and Figure 4 show the results of the comparisons with the IRIS action templates. Using a series of one-tailed, 2x2 Chi-Squared tests without Yates' corrections, the IRIS templates never significantly outperformed the random baseline. Some templates, namely 1, 5, 6, and 11, performed poorly against all four of the other fragment creation processes. This is significant because they were included in the follow up experiment and also performed poorly, where as

Table 1: p values for IRIS action templates versus the five other creation processes in the initial experiment. A one-tailed Chi-Squared test was performed versus the random baseline, and a two-tailed Chi-Squared test was performed versus the others. A dash indicates that more than 1/3 of the responses rated neither as suspenseful.

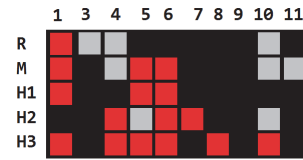|	Rand	MINSTREL	Hum 1	Hum 2	Hum 3
T 1	**0.03**	**0.03**	**0.05**	0.18	**0.03**
T 3	-	0.09	0.06	0.37	0.09
T 4	-	-	0.11	**0.02**	**0.03**
T 5	0.11	**0.03**	**0.01**	-	**0.001**
T 6	0.3	**0.001**	**0.05**	**0.001**	**0.03**
T 7	0.24	0.87	0.55	**0.05**	0.3
T 8	0.36	1	0.11	0.06	**0.05**
T 9	0.18	0.66	0.51	0.06	0.29
T 10	-	-	0.54	-	**0.002**
T 11	0.44	-	0.73	0.57	0.43

Figure 4: A visualization of the comparison between the IRIS action templates and the three other creation processes in the initial experiment. Red indicates the the column was rated higher and black that they were not significantly different, both at the p = 0.05 level. Gray indicates that more than 1/3 of the responses rated neither as suspenseful.

the templates with marginal results in the initial experiment performed much better in the follow up experiment. This suggests inter-experiment validity because the templates that did poorly in the first also did poorly in the second. If they had seen a significant improvement, it would have been less likely that the sentence creation in the initial experiment was confounding the performance of the IRIS action templates.

The other fragment creation processes were also compared to the random baseline. The null hypothesis for these comparisons was that they would perform comparably to random. The alternative hypothesis was that they would perform better than random. A one-tailed, 2x2 Chi-Squared test without Yates' correction was performed for human authors 2 ($\chi^2(1)$ = 0.34, p = 0.28) and 3 ($\chi^2(1)$ = 3.00, p = 0.06). Human author 3 seemed to perform better than random, but human author 2 did not. For human author 1 and MINSTREL, the participants reported more counts of preferring random then these methods, so a two-tailed, 2x2 Chi-Squared test was performed for human author 1 ($\chi^2(1)$ = 4.06, p = 0.04) and MINSTREL ($\chi^2(1)$ = 0.85, p = 0.89). This indicates that these methods did not perform better than random; human author 1 performed worse than random.

Initial Experimental Discussion

The results of of this initial experiment showed that all creation processes, with the exception of one of the human authors, were rated similarly with regard to suspense. Also, few of the fragments were rated as unsuspenseful. It was hypothesized that something other than the creation process was introducing suspense, *i.e.*, the sentences themselves. In the follow-up experiment, the story actions were constructed in a template-based way instead of the sentences of the first experiment, to mitigate the effect of discourse-level suspense being introduced into the sentences themselves.

Another interesting result was that about halfway through the initial experiment, it was observed that the one sentence fragment stories were rated significantly lower than fragments with two or more sentences. A one-tailed 2x2 Chi-squared test testing if the one sentence fragment were rated lower than those of two or more was significant ($\chi^2(1)$ = 38.05, p <0.0001). There was no significant difference among fragments with regards to length as long as they had two or more sentences. The initial set of IRIS templates had several one sentence templates. These were removed from the follow-up experiment (described below).

Finally, the participants were given 30 pair-wise comparisons to make among the different story fragments. There was some concern that they may experience fatigue partway through the experiment and begin to answer the questions differently. Pair-wise 2x2 two-tailed Chi-squared tests were run on each pairing of the 30 questions. No significant difference was found for the responses as the survey progresses.

Follow Up Experimental Design

For the follow-up experiment, the domain and story actions were changed but the design and hypotheses remained the same as in the initial experiment. All of the different creation methods being tested produced story fragments using the same fragment creation grammar. An excerpt of the grammar is shown in Figure 5. The domain was a murder mystery. In this story, there is a detective that is trying to figure out who killed the host of a dinner party. This is the goal that all of the story fragments will try to create suspense around.

To prevent the possible introduction of the discourse-level suspense that proved problematic in the initial experiment, a grammar was used for the creation of story fragments. Fragment length was limited to range between 2 and 6 sentences in order to prevent readers from potential distraction regarding the suspenseful goals. Additionally, the same person or place was not allowed to be used for both parts of a fragment. For example "The maid attacks the maid" was not allowed.

A human author with creative writing experience was recruited to produce suspenseful story fragments. He was given the creation grammar, the rules described in the previous paragraph, and told to create as many suspenseful story fragments around the goal of the detective trying to find the killer as he felt he could generate. He produced seven story fragments between lengths four and six sentences.

Next, a set of story fragments were created using the MINSTREL method of creating suspense (see Section 2). Since IRIS is designed to create suspense at the story-action level and not the discourse level, the step where emotion is

interjected at the discourse level was omitted. An example MINSTREL fragment was "The maid attacked the detective. The detective tried to move to the attic but failed."

Then, the story fragments from the IRIS action formulas shown in Figure 3 were created. There were multiple ways to create fragments from each of the action formulae, so all of the combinations were generated. There is some subtlety in templates 3, 7, & 8, and 4, 9, & 10. In the first three, an action is performed without the main character's knowledge, and in the second three an action is performed without that character being present. However, the suspense is being generated at the story-action level and not the discourse level, so the templates do not make any commitment as to how this action is conveyed to the audience. Since the participants will be reading the story as text and not LISP-style action syntax, a thin but necessary discourse is introduced. These are the template-based "Modifiers" described in Figure 5.

For the logically and temporally consistent baseline, a program was created that generated random story fragments following the story creation rules. The actual story fragments were generated at run time during each participant's experimental session.

The study was conducted online in HTML with JavaScript to help handle the selection and display of the story fragments. Participants were recruited by snowball sampling in email and in person, and through undergraduate and graduate computer science classes. For students recruited through courses, participation fulfilled an experimental course requirement. The participants were first given a link to the consent form. Then, they were given a demographic survey asking their age and gender. In the demographic survey and the comparisons, the participants always had the option to leave any question unanswered. Next, they were given the background information about the domain of the story.

Actions	
<person> observes <condition>.	
<person1> attacks <person2>.	
<person> moves to <place>.	
The detective tries to arrest <person>.	
Conditions	
that <person> is dead	that <person> is at <place>
Modifiers	
(...) but fails	(Unbeknownst to the detective) ...
People	**Places**
The detective	The dining room
The maid	The bedroom

Figure 5: Excerpt of the story fragment creation grammar.

The participants were then given a series of sentence fragment pairs. They were asked to read each fragment and indicate which was more suspenseful. They were given four options: Story Fragment A was suspenseful and Story Fragment B was not, Story Fragment B was suspenseful and Story Fragment A was not, Both were suspenseful, or Neither were suspenseful. After selecting their choice and hitting the Next button, they were shown another pairing. A screenshot of the survey page is shown in Figure 6.
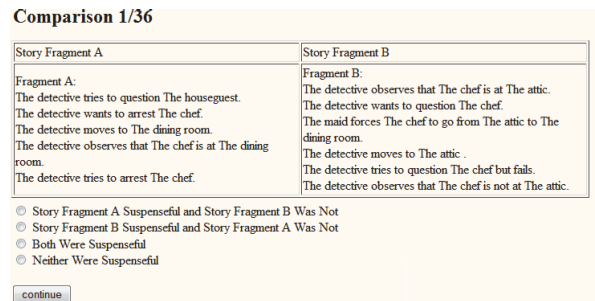

Figure 6: A screenshot of the survey, where participants read and compared the suspensefulness of two story fragments.

There were 36 pair-wise combinations: 11 IRIS action templates x 3 other story creation methods + the 3 combinations of these 3 methods. At run time, a random fragment from the possible ones of the given approach was selected.

The same hypotheses were tested as in the previous experiment.

Follow-Up Experimental Results

83 people completed the demographic survey and were presented with the experiment, with an average age of 22.1 +/- 2.7. There were 9 women and 74 men who participated in the experiment. There were too few women to separate them out and compare their responses to those of the men, so this is a potential threat to external validity. A sizable portion of the participants were computer science undergraduates, which accounts for the low average age. To evaluate if the IRIS templates would be rated comparably to the human authors and if the templates would be rated comparably to MINSTREL, we used the two-tailed 2x2 Chi-square test without Yates' correction to compare the participants' response to a null hypothesis of equal responses for both approaches. The responses for "Both were suspenseful" were added to the counts for believing one method or the other was more suspenseful. For example, 32 people reported that they found template 7 more suspenseful than MINSTREL, seven found MINSTREL more suspenseful than template 7, 12 reported them both suspenseful, and 12 reported neither suspenseful. So the 12 that reported both were suspenseful were added to the counts for each of the two approaches, bringing the count up to 44 for template 7 and 19 for MINSTREL. This was compared with a null hypothesis of 31 for each, which is half of the responses going to each approach (Chi-square tests can't use decimal numbers). If there was no difference in the two methods, then the reported counts would not differ significantly from the expected results. In this case, $\chi^2(1)$ = 5.13, p = 0.02, indicating that the participants report template 7 more suspenseful than MINSTREL.

To evaluate Hypothesis 2, that the IRIS templates would be rated higher than random, and Hypothesis 3, that the human authors would be rated higher than random, we used the one-tailed 2x2 Chi-square test without Yates' correction to compare the participants' response to a null hypothesis of equal responses for both approaches. Here the process was

Table 2: p values for IRIS action templates versus the three other creation processes in the follow-up experiment. A one-tailed Chi-Squared test was performed versus the random baseline, and a two-tailed Chi-Squared test was performed versus the other two. A dash indicates that more than 1/3 of the responses rated neither as suspenseful.

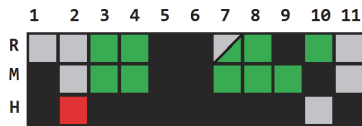|	Rand	MINSTREL	Human
T 1	-	0.46	0.06
T 2	-	-	**0.0001**
T 3	**0.001**	**0.01**	0.47
T 4	**0.003**	**0.05**	0.92
T 5	0.35	0.44	0.12
T 6	0.34	0.52	0.09
T 7	**0.002**	**0.02**	0.10
T 8	**0.001**	**0.05**	0.39
T 9	0.13	**0.04**	0.52
T 10	**0.04**	0.17	-
T 11	-	-	0.51

Figure 7: A visualization of the comparison between the IRIS action templates and the three other creation processes in the follow-up experiment. Red indicates the column was rated higher, green that the row was rated higher, black that they were not significantly different, all at the p = 0.05 level. Gray indicates that more than 1/3 of the responses rated neither as suspenseful. One comparison was right at this 1/3 but had a significant p value.

similar to before, except that the responses "Both were suspenseful" were discarded since we wanted to measure if one performed higher, not just comparable.

In all cases, if more than 1/3 of the responses were "Neither were suspenseful," then the above tests were not performed. This is a threshold we decided on to indicate that the participants found both approaches to lack suspense.

One-tailed 2x2 Chi-Squared tests were performed to measure if MINSTREL and the human author performed significantly better than random. The test for MINSTREL failed to reject the null hypothesis of performing comparably to the random base line, ($\chi^2(1) = 0.01$, p = 0.45), but the test for the human author rejects the null hypothesis of performing comparably to the random, ($\chi^2(1) = 2.33$, p = 0.06).

Follow-Up Experimental Discussion

Hypothesis 3, that the human author would perform better than random, was supported. This is important because a human author is often considered the ideal that story generation systems strive towards, and it needs to be shown that the human author performed well enough to make useful claims when comparing a generation system to the human author.

Hypothesis 4, that the IRIS templates would perform at least as well as MINSTREL, was also supported. The IRIS fragments performed much better than MINSTREL, and MINSTREL did not perform significantly differently than random. This could be due to the fact that MINSTREL requires some discourse level suspense that was purposively minimized in the experiment.

For six of the 11 IRIS templates, Hypotheses 1 and 2 were proven as well. The exceptions were templates 1, 2, 5, 6, and 11. These templates will be removed from the IRIS template bank in future work. The most successful templates were those where the audience had knowledge about the protagonist's imminent plan failure and the protagonist did not. This makes intuitive sense and matches the system's definition of suspense which involves anxiety and uncertainty around the success of an outcome from the audience's perspective.

These results show that it is possible to create localized, story-action level moments of suspense in a story using a small set of action templates comparable to a human author selecting from the same set of story actions.

Future Work

With the suspense action templates having been experimentally verified, they will be incorporated into the IRIS narrative planning system. The action templates will be combined with a cognitive model of the protagonist along with a model of intention revision to produce a complete narrative planner.

The action templates also have value independent of the IRIS system. It has been shown that suspense, comparable to a human author selecting from the same set of actions, can be created by following a small set of action templates. These can be incorporated into a range of narrative generation systems, so long as the action templates can be built into those systems. The action templates can also help guide human authors themselves. Authors can use the templates as building blocks for suspense action sequences which could then be made even more suspenseful in different ways by the addition of discourse level suspense. Alternatively, discourse could augment another narrative phenomenon to create templates that are instead mysterious or surprising.

Conclusion

The IRIS action templates were intended to create localized moments of suspense at the story-action level in a narrative. They were evaluated against several other suspenseful fragment creation processes. Participants read and made pairwise comparisons rating the suspense of these story fragments. It was found that human authors and most of the IRIS action templates, particularly those involving a disparity of knowledge between the audience and the protagonist, were rated significantly more suspenseful than the random baseline. Additionally, the IRIS action templates were rated comparably to the human author. This validates the IRIS action templates as a viable means of creating story-action level suspense in narrative generation.

References

Cheong, Y., and Young, R. 2008. Narrative generation for suspense: Modeling and evaluation. *Interactive Storytelling* 144–155.

Fendt, M. W. 2010. Dynamic social planning and intention revision in generative story planning. In *Proceedings of the Fifth International Conference on the Foundations of Digital Games*, 254–255. ACM.

Gerrig, R. J., and Bernardo, A. B. 1994. Readers as problem-solvers in the experience of suspense. *Poetics* 22(6):459–472.

Riedl, M. O., and Young, R. M. 2005. Open-world planning for story generation. In *Proceedings of the international joint conference on artificial intelligence (ijcai)*, 1719–1720.

Rodell, M. F. 1943. *Mystery fiction: theory and technique*. Duell, Sloan and Pearce.

Turner, S. 1994. *The creative process: A computer model of storytelling and creativity*. Lawrence Erlbaum.

Vorderer, P. 1996. Toward a psychological theory of suspense. *Suspense: Conceptualizations, theoretical analyses, and empirical explorations* 233:254.

Wulff, H. J. 1996. *Suspense and the influence of cataphora on viewers' expectations*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Zillmann, D. 1996. The psychology of suspense in dramatic exposition. *Suspense: Conceptualizations, theoretical analyses, and empirical explorations* 199–231.