Toward a Computational Model of Affective Responses to Stories for Augmenting Narrative Generation

Brian O'Neill

School of Interactive Computing, Georgia Institute of Technology
boneill@cc.gatech.edu

Abstract. Current approaches to story generation do not utilize models of human affect to create stories with dramatic arc, suspense, and surprise. This paper describes current and future work towards computational models of affective responses to stories for the purpose of augmenting computational story generators. I propose two cognitively plausible models of suspense and surprise responses to stories. I also propose methods for evaluating these models by comparing them to actual human responses to stories. Finally, I propose the implementation of these models as a heuristic in a search-based story generation system. By using these models as a heuristic, the story generation system will favor stories that are more likely to produce affective responses from human readers.

Keywords: Affective computing, narrative cognition, computational models, dramatic arc, suspense, surprise

1 Introduction

Narrative as entertainment, in the form of oral, written, or visual storytelling, plays a central role in many forms of entertainment media, including novels, movies, television, and theatre. One of the reasons for the prevalence of storytelling in human culture may be due to the way in which narrative is a cognitive tool for situated understanding [1, 2]. This narrative intelligence is central in the cognitive processes that we employ across a range of experiences, from entertainment contexts to active learning. Expert storytellers who craft narratives for entertainment – films, novels, games, etc. – often structure their narratives to elicit an emotional response from the viewer, reader, or player. The concept of the dramatic arc, identified by Aristotle [3], is one common pattern of emotional impact on an audience.

The construction of novel quality stories is a challenging task, even for humans. For more than 30 years, computer scientists have been trying to answer the question of whether, and how, intelligent computational systems can create stories from scratch. To date, story generation systems have been unreliable when it comes to creating novel and aesthetically pleasing stories with dramatic structure. Zagalo et al. [4] argue for the use of dramatic arc and intelligent emotion detection in story generation and storytelling systems. However, we are not aware of any systems that have adequately adopted this approach to story generation. Simply put, story

generation systems do not have sufficient understanding of story aesthetics nor how story structure affects emotional change in an audience.

The primary objective of my research is to develop an intelligent system that incorporates models of dramatic arc and human affective responses to suspense and surprise for the purpose of story generation. There are many ways to produce dramatic arc in a story [5]. Two related approaches to producing dramatic arc are to make stories suspenseful or surprising; Abbott describes suspense and surprise as the two things that "give narrative its life" [6]. Despite the importance and prevalence of suspense as a storytelling tool, there has been little investigation of computational techniques for generating or understanding suspense. This paper describes work towards a computational model of affective responses to stories, focusing on suspense and surprise, and a story generation system that makes use of those models. The following sections describe related work and the objectives of this research. Sections 4-5 describe current progress on this work and planned future work, respectively.

2 Related Work

Suspense occurs in an audience – the reader or watcher of a narrative – when the audience perceives that a protagonist is faced with the possibility of an undesirable outcome. Gerrig and Bernardo suggest that one method used by authors to make readers feel suspense is to reduce the quantity and/or quality of plans available to the protagonist for avoiding an undesirable outcome [7]. They suggest that readers act as problem-solvers on behalf of the protagonist and when readers can only devise low-quality plans, or struggle to come up with any plans for a hero to escape the predicament, the perception of suspense will increase. In these studies, they found that readers reported higher suspense levels when story excerpts suggested potential escapes and then quickly eliminated them, thus reducing the quantity of available plans for the protagonist.

Branigan [5] suggests that suspense and surprise are the result of knowledge disparities between the audience and the characters. An author creates suspense by providing the audience with more knowledge than the characters, particularly about the possibility of undesirable outcomes. Conversely, surprise is created when the characters possess more knowledge than the audience. Suspense has also been described as a lack of closure within a narrative [6]; authors manipulate readers by appearing to satisfy the need for closure, only to take it away. Abbott describes successful narratives as chains of suspense and surprise.

To this point, story generation systems have been unreliable at creating novel stories with dramatic arc. However, there has been progress at generating surprising and suspenseful discourses for existing stories. Suspenser [8] computationally attempts to find a suspenseful telling of an existing story. Suspenser tries excluding different sets of events from the discourse to maximize a suspense rating. The level of suspense of any telling of a story is measured by generating all possible plans a protagonist might have and taking the ratio of failed plans to successful plans. Prevoyant [9] uses a computational model of flashback and foreshadowing to produce a reordered version of a story intended to elicit feelings of surprise in human readers.

Prevoyant rearranges the events of a given story in order to produce surprise from outcomes that are unexpected by readers. Prevoyant does not measure the level of unexpectedness; rather, unexpectedness is defined as the reader's inability to find a plan explaining the outcome without the events of the flashback. Each of these systems receives the complete story as input and modifies the discourse of those stories – which events of the story are told, and in what order – to produce a suspenseful or surprising ordering of events. In my research, I do not seek to modify the discourse of existing stories; rather, I want to read the story incrementally, and identify the aspects of the story that produce suspense or surprise responses. These will be used to create a model of suspense and surprise that can then be applied as a heuristic for the generation of stories that produce affective responses.

Story generation systems solve the problem of finding a sequence of events that can be told to a human audience as a story. Some story generation systems use non-emotional aesthetic qualities, such as novelty [10] or character believability [11]. A number of story generation and interactive narrative systems apply tension ratings; however these ratings are hard-coded into discrete events as absolute ratings. None of these tension ratings are based on models of affective response. The MEXICA story generation system [10] models reader tension as a measure of how satisfactory the state of the story is. Certain events in MEXICA define an increase or decrease in tension as an effect of the event. The amount of the change in tension for a given event is defined in advance by the user of the system. MEXICA tracks the tension over the course of the story and compares the changes in tension to other stories that it knows. Porteous et al. [12] describe an interactive storytelling system that allows users to control the level of tension in the story. The system produces a pacing and ordering of events that best fits the curve representing the provided tension levels. The interactive drama, Façade [13], tracks tension as a factor in managing the interactive narrative. The Façade drama manager has an ideal tension curve and probabilistically changes the tension in the narrative to try to match the ideal curve. Each event in the story has a number of possible presentations based on character affinity and the tension level. The drama manager selects a presentation based on how well it fits the scene so far and the ideal tension level at that point of the story. Rather than hard-coding tension ratings, I argue that a better approach would be to base these ratings on models of affective response.

3   Research Objectives

The objective of this research is to address two primary questions related to computational modeling of aesthetics and affect in narrative:

- How do we computationally model affective response to stories?
- How can we utilize computational models of affective response to stories to automatically generate more suspenseful or more surprising stories?

As described above, current approaches to story generation do not use affective response as a means of creating new stories. Many of these story generation systems
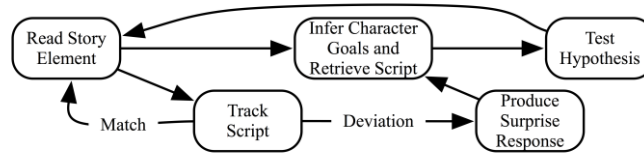
Figure 1. This flowchart represents the process used by the Surprise Model to read a story, infer character goals, and track scripts in order to recognize surprise responses.

also lack a model of dramatic arc. It is my belief that we can generate better stories by incorporating each of these elements into the story generation process. In my efforts to address these research questions, I plan to implement two models of affective response to stories, emphasizing surprise and suspense. I will then incorporate these models into a story generation system – effectively using them as a heuristic to guide the story generation process. Evaluation of the story generation system will involve asking readers to compare stories produced by this system to a comparable story generator that lacks models of dramatic arc and affective response.

4 Current Progress

My current progress towards the development of a model of affective response is made up of two cognitively-plausible computational models, one each for the generation of surprise and suspense responses to narratives. At this stage, neither model has been fully implemented. The Surprise Model is based on the construction of a possible reader model as the story is being read, focusing on the superordinate goals of the characters in the story [14]. The Suspense Model tracks the goals and plans of the protagonist, identifies possible failures in each, and assesses the likelihood of escape from failure states [15].

4.1 Surprise Model

The Surprise Model, based on an earlier model of computational narrative comprehension [14], constructs one possible reader mental model of the story incrementally, as it is being read. Authors and directors can achieve audience surprise in several ways. One such approach is by providing the audience with less information that the characters in the story have [5]. Using this observation, I focus on surprise that originates from a character doing something unexpected. The Surprise Model focuses on character goals because (a) they are easily inferred by readers, and (b) actions in stories are typically driven by character goals. Readers feel surprise when the predictions they make about the story, such as inferring characters' superordinate goals, are incorrect. Figure 1 shows the process used by the Surprise Model to identify surprise responses.
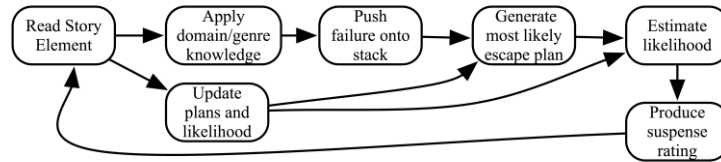
Figure 2. This flowchart represents the process used by the Suspense Model to move from reading part of a story to producing a suspense rating.

A story is provided to the Surprise Model as input, and story elements are read one at a time. I do not address any aspect of natural language understanding or computer vision necessary to literally read or view a creative artifact such as a story or film. Rather it reads an annotated version of the story or film script, where an annotation contains relevant information about the scene, characters, and non-diegetic information that may be used to elicit an affective response from the audience.

The search for superordinate character goals drives the process because of the importance of goals in story comprehension and sense-making [16]. The Surprise Model, acting as a surrogate for a human audience, uses several strategies to hypothesize the superordinate goal for each character that is actively engaged in the current event. These strategies prefer goals that have been explicitly stated in the story or were inferred in previous iterations. In the process of inferring character goals, the Surprise Model may also retrieve a script that expresses likely future actions. The Surprise Model tests the hypothesized goal and script using a narrative planner, which creates a narrative that links the events of the story with the hypothesized character goals. If the planner cannot form a narrative that links the events and hypothesized goals, then the Surprise Model uses a different strategy to identify a goal.

Once a goal and script have been retrieved, the Surprise Model tracks that script as the story continues. When a character deviates from the hypothesized goal or the retrieved script, the Surprise Model re-tests the goal and script using the narrative planner. Surprise responses can occur when the inferred goal was incorrect or when characters deviate from the expected plan to carry out the inferred goals. When this occurs, the Surprise Model infers a new goal based on the new information in the story and retrieves a new script to predict the characters' actions.

4.2 Suspense Model

The Suspense Model, proposed in [15], is based on the correlation between the perceived likelihood of a protagonist's failure and the amount of suspense reported by the audience. The Suspense Model reads the story incrementally, attempting to predict failures in the protagonist's goals and plans. The likelihood of averting the predicted failures is used to compute a suspense rating. The flowchart in Figure 2 shows the complete suspense process. The key points of the model are highlighted below.

The Suspense Model reads an annotated story, one element at a time. After reading an element, the next step is the application of domain and genre knowledge. The

Suspense Model contains scripts and schemas pertaining to common stories and genre techniques in order to simulate the domain and genre knowledge that authors and directors regularly expect their audiences to have. These scripts and schemas are also used to identify the goals and/or plans of the protagonist. Once the protagonist's plan has been identified, the system searches for a potential failure in that plan.

The Suspense Model uses its knowledge of the story, the protagonist's goals and plans, and the potential failure to find an *escape plan*, a plan that allows the protagonist to avoid the undesirable outcome while still completing his original plan. We generate an escape plan for the purpose of identifying the likelihood of the protagonist's escape, which is correlated with the suspense felt by a human audience. As events that are more easily retrieved from memory are perceived by humans to be more likely [17], the system presumes that the first escape-plan is the most likely to succeed. The likelihood of the plan is calculated as a function of the state of the world, the costs of the actions in the plan, and the time available before the failure can no longer be averted.

Finally, the Suspense Model uses this likelihood to calculate a suspense rating. Recall that Gerrig and Bernardo [7] found that reducing the quantity or quality of plans available to the protagonist led to readers reporting higher levels of suspense. Plans with lower likelihoods of succeeding can be seen as being of poorer quality. Thus, as likelihood of escape increases, the suspense rating is expected to decrease. The calculation of suspense rating also factors in the severity (negative utility) of the failures and the audience's affinity for the characters in question.

Once the Suspense Model has calculated a suspense rating, it continues to the next event in the story. As events are added, the model (a) updates the suspense rating, (b) finds a new escape-plan and recalculates suspense, or (c) notes the aversion of the potential failure and moves on to identifying or resolving other potential failures.

5 Future Work

The remaining work on the Surprise and Suspense Models is implementation and evaluation. I plan to (a) assist in the knowledge engineering problem for each model, (b) evaluate the ratings provided by the models with ratings provided by a human audience by correlating with the suspense and surprise levels reported by actual human audiences, and (c) demonstrate the effective use of these models as heuristics for generating stories with dramatic arc.

5.1 Knowledge Engineering

The Surprise and Suspense Models both require a library of scripts and schemas representing domain and genre knowledge. To avoid researcher bias, I plan to acquire these scripts through experimental procedures. Participants will be asked to provide fragments of stories, delineating the events that occur in these fragments and identifying the superordinate goals of the characters in these situations. These responses will be compiled into script-like structures, based on the plan networks described by Orkin [18]. Compiling the responses into scripts will be automated if

possible, but may not be necessary for small-scale experimentation with the models.

5.2 Evaluation of Models

Once the Surprise and Suspense Models have been implemented, I will conduct evaluations of the ratings that these systems produce, in order to be certain that the ratings produced by the models reflect actual human affective responses. In this study, I plan to ask human subjects to watch a movie and have them provide ratings of their suspense and surprise levels as they watch. These ratings will be compared to the ratings produced by the Surprise and Suspense Models. It is not necessary for the numerical ratings to match precisely between humans and the affective response models. Rather, the evaluation will look to see that the relative increase and decrease in suspense levels are comparable between humans and these models, and that the relative suspense levels at different points of the movies are also comparable.

5.3 Applying Affective Models to Story Generation

Current attempts at story generation have not been able to reliably produce aesthetically pleasing stories with dramatic structure without manual, ad-hoc encoding of tension values for states and actions. I argue that this is the result of story generation systems not using cognitively plausible models of affective response as part of the generation process. I propose to integrate the Surprise and Suspense Models into a search-based narrative generation system to show that the models can improve the aesthetic quality of computer-generated stories. Search-based narrative generation has been a popular approach, in part, because of the many correlations between AI plans and narrative structures [11]. A search-based story generator solves the problem of finding a sequence of story events that achieves a set of causal, structural, and aesthetic requirements. The search process is guided by a heuristic function that may be based on a number of factors, possibly including length, believability, and a number of ad-hoc rules about the story. By incorporating the Surprise and Suspense Models into the heuristic, a search-based narrative generator will, in a principled manner, favor narrative structures that lead the reader to affective responses.

To evaluate the narrative generation system, I aim to show that the inclusion of the Surprise and Suspense Models can result in greater affective responses and quality judgments in human readers than in an equivalent narrative generation system that does not use my models. I propose the following experimental design. Two story generators will be developed; the control and experimental systems will be identical except for the inclusion of the affective response models in the experimental system. Both story generators will be provided with the first half of a story. This story will be handcrafted and empirically evaluated to trigger concern for a protagonist. By holding the first half of the story constant, we can avoid unfair comparisons on the variables that we care about. Each system will complete the story and human study participants will be asked to rate stories generated by both systems, indicating affective response levels as well as subjective indications of overall story quality. We hypothesize that

the stories generated by the experimental system, implementing the affective response models, will receive higher ratings from the study participants. If the within-subjects ratings of stories produced by the model-guided story generator are better than those created by the control system, then this work will represent a valuable contribution to a number of fields, including affective computing, particularly in the context of stories, as well as computational creativity and aesthetics.

References

1. Bruner, J.: The Narrative Construction of Reality. Critical Inquiry. 18, 1--21 (1991).
2. Gerrig, R.J.: Experiencing Narrative Worlds: On the Psychological Activities of Reading. Yale University Press, New Haven (1993).
3. Aristotle: The Poetics (T. Buckley trans.). Prometheus Books, Buffalo, NY (1992). (Original work published 350 B.C.E.).
4. Zagalo, N., Barker, A., Branco, V.: Story reaction structures to emotion detection. In: Proceedings of the 1st ACM Workshop of Story Representation, Mechanism, and Context. p. 33--38. ACM Press, New York (2004).
5. Branigan, E.: Narrative Comprehension and Film. Routledge, New York (1992).
6. Abbott, H.P.: The Cambridge Introduction to Narrative. Cambridge University Press, Cambridge (2008).
7. Gerrig, R.J., Bernardo, A.B.I.: Readers as Problem-Solvers in the Experience of Suspense. Poetics. 22, 459--472 (1994).
8. Cheong, Y.-G.: A Computational Model of Narrative Generation for Suspense. Doctoral dissertation, North Carolina State University (2007).
9. Bae, B.-C., Young, R.M.: A Use of Flashback and Foreshadowing for Surprise Arousal in Narrative Using a Plan-Based Approach. Proceedings of the International Conference on Interactive Digital Storytelling. LNCS, vol. 5334, pp. 156-167. Springer, Heidelberg (2008).
10. Pérez y Pérez, R., Sharples, M.: MEXICA: A Computational Model of a Cognitive Account of Creative Writing. Journal of Experimental and Theoretical Artificial Intelligence. 13, 119--139 (2001).
11. Riedl, M.O., Young, R.M.: Narrative Planning: Balancing Plot and Character. Journal of Artificial Intelligence Research. 39, 217--268 (2010).
12. Porteous, J., Teutenberg, J., Pizzi, D., Cavazza, M.: Visual Programming of Plan Dynamics using Constraints and Landmarks. In: Proceedings of the 21st International Conference on Automated Planning and Scheduling. p. 186--193. AAAI Press, Freiburg, Germany (2011).
13. Mateas, M.: Interactive Drama, Art and Artificial Intelligence. Doctoral dissertation, Carnegie Mellon University (2002).
14. O'Neill, B., Riedl, M.: Simulating the Everyday Creativity of Readers. Proceedings of the Second International Conference on Computational Creativity. p. 153--158. AAAI Press, Mexico City, Mexico (2011).
15. O'Neill, B., Riedl, M.: Toward a Computational Framework of Suspense and Dramatic Arc. In: Proceedings of the 4th International Conference on Affective Computing and Intelligent Interaction (2011) (in press).
16. Graesser, A.C., Singer, M., Trabasso, T.: Constructing Inferences During Narrative Text Comprehension. Psychological Review. 101, 371--395 (1994).
17. MacLeod, C., Campbell, L.: Memory Accessibility and Probability Judgments: An Experimental Evaluation of the Availability Heuristic. Journal of Personality and Social Psychology. 63, 890--902 (1992).
18. Orkin, J.D.: Learning Plan Networks in Conversational Video Games. Masters of Science thesis, Massachusetts Institute of Technology (2007).