Proceedings, The Eleventh AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE-15)

Evaluating the Pairwise Event Salience Hypothesis in *Indexter*

Christopher Kives and **Stephen G. Ware**
crkives@uno.edu, sgware@uno.edu
Narrative Intelligence Lab
University of New Orleans
New Orleans, LA, USA

Lewis J. Baker
lewis.j.baker@vanderbilt.edu
Thinking and Seeing Lab
Vanderbilt University
Nashville, TN, USA

Abstract

Indexter is a plan-based computational model of narrative discourse which leverages cognitive scientific theories of how events are stored in memory during online comprehension. These discourse models are valuable for static and interactive narrative generation systems because they allow the author to reason about the audience's understanding and attention as they experience a story. A pair of *Indexter* events can share up to five indices: *protagonist*, *time*, *space*, *causality*, and *intentionality*. We present the first in a planned series of evaluations that will explore increasingly nuanced method of using these indices to predict salience. The Pairwise Event Salience Hypothesis states that when a past event shares one or more indices with the most recently narrated event, that past event is more salient than one which shares no indices with the most recently narrated event. A crowd-sourced ($n = 200$) study of 24 short text stories that control for content, text, and length supports this hypothesis. While this is encouraging, we believe it also motivates the development of a richer model that accounts for intervening events, narrative complexity, and episodic memory decay.

1 Introduction

Narratives are a form of communication between the author and the audience. When designing a narrative, skilled authors pay careful attention to how it will be experienced in order to affect the audience's comprehension of events in the narrative's past, present, and future. One key aspect of online narrative comprehension is *salience*, the ease with which the audience can recall a past event.

Reasoning about salience is not only valuable for analyzing static stories but also for generating dynamic ones. Interactive narrative systems that reason about salience can control which events the audience remembers most readily. This has applications for generating discourse phenomena such as suspense and surprise. It may also have an important influence on how the audience reasons about the future of the narrative.

A wealth of empirically validated cognitive science research (see survey by Zwaan and Radvansky (1998)) has demonstrated key features that the mind uses to store and retrieve events in short term memory when experiencing a story. Previous work (Cardona-Rivera et al. 2012) integrated these indices into a plan-based model of narrative, named *Indexter*, and hypothesized that they could be used to measure the salience of past events during comprehension. Events in an *Indexter* plan can share up to five narrative situation indices with one another: *protagonist*, *time*, *space*, *causality*, and *intentionality*. The presence or absence of these indices in each event of the discourse should influence the salience of past events.

There are several ways this model might be used to measure salience. This paper presents the first in a planned series of evaluations by which we hope to discover the link between *Indexter's* indices and salience by considering progressively more complex models. We begin with the simplest hypothesis, which we call the Pairwise Event Salience Hypothesis: when a past event shares one or more indices with the most recently narrated event, that past event is more salient than one which shares no indices with the most recently narrated event.

Participants read short stories from four different domains. At a given point in each story, they were interrupted and asked to recall details from a past event which shared a certain index with the most recent event. The reader's reaction time is used as a proxy to measure salience. A crowd-sourced study ($n = 200$) rejects the null hypotheses.

This evaluation is important for several reasons. While each index has been studied individually and some subsets have been studied together, to our knowledge no study has ever considered the relative contributions of all five to narrative salience in the same experimental context. Also, some indices may have stronger effects than others, and this experimental design has the potential to reveal that order of importance, but this study was insufficiently powered to do so. While the results of this experiment are encouraging, we believe they also motivate the development of a richer theory of how *Indexter's* indices are used to predict salience.

2 Related Work

Narratologists often divide a narrative into *story* (people, places, things, and events) and *discourse* (how the story is told) (Bal 1997). The AI *planning* formalism provides both a rich knowledge representation and a suite of algorithms for generating and adapting both of these aspects of narrative.

Indexter is primarily concerned with discourse because it represents parts of the audience's mental model during the comprehension process. It specifies which events from the story should be presented and in what order. Similar plan-based models have been applied to other discourse phenomena, such as suspense (Cheong and Young 2008), surprise (Bae and Young 2014), and cinematic composition (Jhala and Young 2010). Numerous plan-based models have been used to reason about story structure and to control interactive stories (see survey by Young et al. (2013)). As with these other models of discourse, *Indexter* can inform story generation as well as discourse generation.

Indexter defines a plan data structure augmented with a cognitive scientific model of narrative comprehension called the event-indexing situation model, or EISM (Zwaan and Radvansky 1998). EISM is the result of decades of empirical research on how audiences store and retrieve narrative information in memory while experiencing a narrative. Zwaan and Radvansky (1998) identify five important dimensions, or indices, of narrative events which have been shown to play a role in narrative comprehension: *protagonist* (who), *time* (when), *space* (where), *causality* (what enabled or impelled), and *intentionality* (why).

A previous study used *Indexter* to predict agency (Cardona-Rivera et al. 2014). When choosing between two alternatives in an interactive hypertext adventure game, players self-reported a higher sense of agency when the perceived next state that would result from making each choice differed from one another in at least one index. This indicates that players reason about the five indices, but does not test the link between recently narrated events and past events.

Specifics about *Indexter* are given in the next section. Various cognitive science studies relevant to our methodology are referenced in Section 4 when we describe our experimental design.

3 The *Indexter* Model

Indexter defines a data structure for representing stories as plans. A pair of events in a story can share up to five dimensions with one another: *protagonist*, *time*, *space*, *causality*, and *intentionality*. This section reproduces very briefly those definitions needed to understand the evaluation described in this paper; for a detailed description of how *Indexter* maps EISM indices to plan structures, see Cardona-Rivera et al. (2012).

A planner is an algorithm which attempts to solve the following problem: given a world in some initial state, a goal, and a set of possible events, find a sequence of those events which achieves the goal. Each event has preconditions which must be true immediately before it is executed and effects which modify the world state. The solution returned by a planner is a sequence of events called a plan.

The kinds of events that can occur are represented by abstract, parameterized templates called operators, as described by STRIPS formalism (Fikes and Nilsson 1972). For example, the domain might define an operator $attack(?attacker, ?victim, ?place, ?time)$. Each term starting with a '?' is a free variable which can be bound to a logical constant corresponding to some specific thing in the story world. The preconditions might be that the attacker and victim are both alive, that both are in the same place at the same time, and that the attacker is armed. The effects might be that the victim is no longer alive. An *Indexter* event is a fully ground instance of such an operator.

Riedl and Young (Riedl and Young 2010) define an extended kind of planning called *intentional planning* which reasons about not only the author's final goal for the story but also each individual character's goals. Intentional planning attempts to ensure that plans express believable character behaviors which are clearly motivated and goal-oriented.

Each operator in an intentional planning domain specifies zero, one, or many of its parameters as being the consenting characters responsible for taking that action. For the $attack$ example, the $?attacker$ is the sole consenting character, because it carries out the action. While the $?victim$ may be a character involved in the action, it need not be a consenting character.

Definition 1. Two events share the *protagonist* index iff they have one or more consenting characters in common.[1]

Each event in an *Indexter* plan must also specify two additional required parameters: the time frame in which it occurs and the location at which it occurs. For example, the $attack$ action might specify that $?time = day1$ and $?place = cave$.

Definition 2. Two events share the *time* index iff their time parameters are the same symbol.

Definition 3. Two events share the *space* index iff their location parameters are the same symbol.

Cognitive science research (Magliano, Miller, and Zwaan 2001; Zacks, Speer, and Reynolds 2009) has demonstrated that time and space can be hierarchically organized in memory. Whether different rooms in the same house count as different locations depends on the discourse. *Indexter* uses a highly simplified representation of these concepts as unique symbols. For this representation to be effective, the discourse must communicate the appropriate level of granularity to the audience. Section 4 describes how this was achieved for the stories used in this study.

One strength of the plan-based models of narrative on which *Indexter* is based is the ability to reason about causal relationships between events. While cognitive scientists have studied several forms of causality (Trabasso and Sperry 1985; Trabasso and Van Den Broek 1985; Zwaan and Radvansky 1998), one in particular is easily available in plans using causal links: the ways in which the effects of earlier events enable later events by establishing their preconditions.

Definition 4. A *causal link* $s \xrightarrow{p} t$ exists from event s to event t for proposition p iff s occurs before t, s has the effect p, t has a precondition p, and no event occurs between s and t which has the effect $\neg p$. We say s is the *causal parent* of t, and that an event's *causal ancestors* are those events in the transitive closure of this relationship.

[1] Here we use the one protagonist per event (as opposed to one per story) definition discussed by Cardona-Rivera et al. (2012).

Definition 5. Two events share the *causality* index iff the earlier event is the causal ancestor of the later event.

Riedl and Young's intentional planning framework organizes events into *frames of commitment* to explain how characters achieve their individual goals. These structures also rely on consenting characters and causal relationships.

Definition 6. Let c be a character and g some goal that character c intends to achieve. Let s be an event with effect g for which c is a consenting character. Two events share the *intentionality* index iff both events have c as a consenting character and both are causal ancestors of s. (Note that s may be one of the events.)

In other words, two events share *intentionality* when both are taken by the same character for the same purpose.

4 Experimental Design

Much of the value of *Indexter* to the interactive narrative community lies in discovering how its operationalization of cognitive scientific theories can be used to generate stories which create specific discourse effects. We desire the simplest effective model of how narrative indices are correlated to salience, so we begin with the simplest possible model, iteratively testing it and incorporating new insights from cognitive science research until a sufficiently powerful model is developed.

A starting point was proposed with the original description of *Indexter* (Cardona-Rivera et al. 2012), which we call the Pairwise Event Salience Hypothesis: when a past event shares one or more indices with the most recently narrated event it is more salient than an event which shares no indices with the most recently narrated event. The authors acknowledge many possible improvements to this hypothesis, and this study attempts to provide insights for that process.

We tested this hypothesis by having audiences read short text stories one event at a time. Some event in the story is designated the *referent*, and some later event is designated the *prompt*. We engineer the content of each story such that the prompt and referent share exactly one or zero indices. After reading the prompt, the participant is interrupted and asked to recall the referent. The speed with which the participant is able to answer, i.e. their reaction time, is used as a proxy to measure salience—the shorter the response time, the more salient the event being recalled.

The Pairwise Event Salience Hypothesis We state five hypotheses, one for each index, *protagonist*, *time*, *space*, *causality*, and *intentionality*. The null hypothesis for an index states that, among subjects who remember the referent event, the reaction time of a subject will not differ significantly when the prompt shares that index with the referent than when it does not share any indices. The alternative hypothesis is that the subject will react faster when the prompt shares that index with the referent than when it does not share any indices.

This experiment is complicated due to the large number of variables that needed to be controlled. In this section, we provide details on various parts of the procedure by identifying those controls and how they are addressed.

	Min	Max	Mean	σ
Length	18	21	18.96	1.04
Prompt Distance	10	18	14.96	2.10
Prompt Gap	3	13	8.25	3.05

Table 1: Summary statistic for various story length properties, given as number of events.

Materials

We designed four fictional story domains in which to test our hypothesis: a zombie apocalypse, a medieval fantasy, a science fiction adventure, and a heist. Six stories were created for each domain, totalling 24 stories.

Story Content For one story in each domain, the referent and prompt shared only the *protagonist* index; for one story they shared only *time*; and so on. There was one exception based on index definitions: the stories for which referent and prompt share *intentionality* also share *protagonist* and *causality*[2]. A sixth story in each domain had a referent and prompt that shared no indices.

We designed the story domains in order to control the content (characters, word choice, places, goals, etc.) across all stories in the domain as much as possible. Some differences had to be introduced between the stories in order to provide a suitable prompt and referent for each index, but wherever possible we have kept the content and order of events consistent. The result, for each domain, is six slightly differing versions of the same story. We designed these domains ourselves because, to our knowledge, no sets of "naturally occurring" stories exist with these highly controlled properties.

Story Length We attempted to control story length in addition to content. Table 1 gives summary statistics for three important length values. Length indicates the total number of events in the story (where the description of the initial state is considered the first event). Prompt Distance indicates the number of events read before the subject's memory is probed. Prompt Gap indicates the number of events between the referent and prompt.

Story Text The text of one story used in the study is given in Figure 1 for reference. The story is divided into 18 events. The first event describes the initial state of the story, including each character's current location and goals. The referent (event 11) and prompt (event 17) are in bold.

Most events are expressed as two sentences. The first is a text representation of a planning operator generated using simple natural language templates (e.g. event 6 was generated from "$?character$ moves into $?place$."). The second is a hand-written piece of flavor text associated with that event. These flavor text sentences are used to recall the referent

[2]Events which share *intentionality* share *protagonist* by definition. While it is technically possible for two events to share *intentionality* without sharing *causality*, we found that this required very convoluted and unnatural stories, so this constraint was not enforced, and instead all events which share *intentionality* also share *causality*.

Story 22

1. This is a story about a thief. Prisoner is in courtroom. Police are in courtroom. Judge is in courtroom. Thief is at thief's home. Thief intends to steal diamonds.
2. Prisoner claims innocence. Silence fills the courtroom.
3. Police submit strong evidence. The sounds of a recording of a crime fills the room.
4. Judge rules guilty. The sound of the gavel emanates from the courtroom.
5. Police escort Prisoner to prison. The cell is cold and damp.
6. Thief moves into the store. It is pitch black.
7. Thief breaks into the vault. Steel lines the walls.
8. Thief takes the diamonds. The diamonds look elegant.
9. Thief moves to thief's home. Flashes of color from the TV flash on the walls.
10. Police move to the store. Officers work loudly.
11. **Police search for evidence. Dim lighting illuminates the room.**
12. Police find weak evidence. A small hair laid on the ground. Police intend to arrest Thief.
13. Police move to thief's home. TV sounds flood the house.
14. Police arrest Thief. She moves calmly.
15. A week passes after the heist. Police escort Thief to the court. Courtroom is noiseless.
16. Thief claims innocence. Her voice is calm.
17. **Police submit the weak evidence. It isn't convincing.**
18. Judge rules not guilty. A smile appears on Thief's face.

Figure 1: One story, from the heist domain, used in the study. The referent and prompt events are in bold. They share only the *intentionality*, *protagonist*, and *causality* indices.

without affecting the audience's mental model (discussed later).

Some events have additional text in special cases. When the time frame changes between two events, such as between events 14 and 15, the second begins with something like "A week passes..." to indicate the change. If an event motivates a new character goal, this is indicated at the end, such as in event 12 of the example, "The police intend to arrest Thief."

To ensure that the text successfully conveys which events do and do not share indices, the three authors acted as raters and tagged each referent and prompt for each of the five dimensions. When raters disagreed, the disagreement was discussed, the stories were modified, and the tagging was performed again until perfect agreement was achieved (Cronbach's $\alpha = 1$).

Segmentation *Indexter* assumes that as the audience experiences a narrative they will segment the events, time, and space into discrete units. It is important that the granularity of the segmentation scheme used when generating the story matches the one used by the audience as closely as possible.

In an attempt to ensure this, all participants played a short interactive choice-based text game at the start of the survey as part of the initial training. The story uses events and text similar to the test stories. One segment of text (i.e. one event expressed as template sentence(s) + flavor text) was displayed at a time. Panels at the bottom of the screen indicated the time frame (e.g. Day 1) and location (e.g. kitchen) of the event currently displayed.

This game is meant to prime the audience on how events, time, and space are segmented. This level of priming is important to ensure that the audience perceives a change in time/space when we intend them to. Previous studies have primed segmentation either through visual aids of spatial arrangement (e.g. Morrow, Bower, and Greenspan (1989)) or through passive viewing of films prior to an experiment (e.g. Zacks, Tversky, and Iyer, (2001)), but in future studies we intend to relax this constraint to improve generalizability.

Procedure

Stories were read online via a web browser. A participant places two fingers from the same hand on two keys (e.g. the "1" and "2" keys on the number row of the keyboard) and the second hand on the spacebar. One event is shown on the screen at a time. The spacebar advances the text on the screen to the next event. After the prompt event is displayed, the participant is interrupted. They are shown text from the referent event and asked to press "1" (yes) if they remember it and "2" (no) if they do not remember. Participants were asked to respond within 2 seconds, but responses over 2 seconds were still recorded. Different keyboard configurations where available for left and right-handed participants.

Measuring reaction time in this fashion is an established means of studying salience in memory-related tasks (e.g. Gillund and Shiffrin (1984), Yonelinas and Jacoby (1994)). Reaction time was measured using Javascript, which is sufficiently accurate for measuring response time across different systems and browsers (Reimers and Stewart 2014).

One significant challenge for the design of this experiment is that the text used to probe the subject's memory may itself influence their mental model. Asking if the subject remembers "Police search for the evidence" may prime their response by mentioning the police character. For this reason, the flavor text of the referent sentence was used to probe their memory (which is why the flavor text was added to the stories). When asked to answer "yes," or "no," for the story in Figure 1, the participant was shown "Dim lighting illuminates the room," which is the flavor text of the referent, event 11. This indirect probing of memory has been used in similar studies (e.g. Brewer and Dupree (1983)).

Before starting the study, participants practiced with the interface on two short stories, repeating them until they were able to answer the prompts correctly in under 2 seconds. Each subject then read four stories, one from each of the four domains. Stories were presented in a random order to control for an ordering effect. Two stories had "yes" answer prompts and two had "no" answer prompts. Subject were told they would only receive compensation if they answered 3 of 4 of the prompts correctly, though in truth participants with lower accuracy were also compensated.

Tests which aim to illuminate the audience's mental model require many participants because little data is collected per person. Interrupting the subject may modify their mental model, so only 1 data point is obtained per user per story. In addition, some prompts must have a "no" answer to stop participants from believing that "yes" is the correct answer for every prompt.

Index	Count	Mean	σ
None	59	1803	688
Protagonist	55	1499	672
Time	52	1554	541
Space	56	1562	426
Causality	52	1623	476
Intentionality	50	1561	490

Table 2: The number of observations per index along with mean response time and standard deviation in milliseconds.

Participants

Due to the large number of data points needed, subjects were recruited via the Amazon Mechanical Turk crowd-sourcing web platform. They were offered a small amount of money for participating (between \$0.50 and \$0.55). Participants were limited to residents of the USA who are 18 years of age or older. 200 participants completed the study on Mechanical Turk, resulting in 800 responses across 24 stories.

We observed considerable variance in reaction times and a surprisingly low accuracy of only 71%. A d' analysis (Macmillan and Creelman 2004) suggests that subjects not sure of the answer were biased toward "yes." ($d' = 1.156$; Hit rate $= 0.820$, false alarm rate $= 0.405$). We suspect these trends are due to the complicated nature of the stories and the high variance in performance of Mechanical Turk workers, which we discuss further in Section 6.

5 Results

Due to the high variance in response time, and because each participant read all 4 stories of the same index, we calculated the mean response time for each index and removed any observations more than 2 standard deviations above or below the index mean (24 of 800 data points).

Because we wish to test how the presence or absence of an index affects *accurate* memory, we also removed all observations for which the correct answer was "no" (400 out of 800 data points) and all observations which were answered incorrectly (234 out of 800 data points). A total of 324 data points remained. These data points are the ones for which a subject correctly answered "yes" in a number of milliseconds within 2 standard deviations of the index mean. A summary of these 324 observations, broken down by index, is given in Table 2.

A one-way repeated measures ANOVA revealed marginally significant differences in response time by index ($F_{5,318} = 2.079$, $p = 0.0677$). For comparing each index to the None condition, we performed paired t-tests, as shown in Table 3. Finally, for comparing indices to one another, we performed paired t-tests using the Benjamini & Hochberg (1995) correction for multiple comparisons, as shown in Table 4.

Regression analyses revealed no significant differences in response time by prompt distance (number of events read since the beginning of the story, $F_{1,320} = 0.003$, $p = 0.96$) or gap (number of events read since the referent, $F_{1,320} = 0.428$, $p = 0.51$), or their interaction ($F_{1,320} = .087$, $p = .77$). This suggests we succeeded in controlling information load between stories. Note that we are only considering correct responses, so accuracy is always 100% and thus not a potential confounding factor.

	None	Protag.	Time	Space	Cause
Protag.	**0.004**				
Time	**0.021**	0.612			
Space	**0.022**	0.553	0.941		
Cause	0.092	0.256	0.535	0.577	
Intent.	**0.026**	0.573	0.951	0.991	0.580

Table 3: p values for paired t-tests of each index, unadjusted. Significant values ($p < 0.05$) are in bold.

	None	Protag.	Time	Space	Cause
Protag.	0.062				
Time	0.096	0.765			
Space	0.096	0.765	0.991		
Cause	0.277	0.640	0.765	0.765	
Intent.	0.096	0.640	0.991	0.991	0.765

Table 4: p values for paired t-tests of each index, adjusted using the Benjamini & Hochberg (1995) correction method.

6 Discussion

We rejected 4 of 5 null hypotheses at the $p < 0.05$ level and all 5 at the $p < 0.1$ level. Participants who accurately remember the referent respond faster when the most recently narrated event shared at least one index with the referent.

At this point we can only conclude that any index in common between prompt and referent benefits memory retrieval compared to having no common indices. This study is insufficiently powered to address whether some indices have a stronger effect than others (as seen in Table 4).

While these results are encouraging, we observed low accuracy on answering questions and high variance in response time. We consider this motivation for the development of a richer model. In this section, we consider potential explanations for our results and discuss how the model can be improved.

Use of Mechanical Turk We found the Mechanical Turk API and documentation very difficult to use. We chose this service because of the high volume of participants needed for the study, but if the volume can be reduced, we hope to replicate this study or conduct future studies in a more traditional lab setting. We are especially interested in suggestions from the memory and narrative research communities on how this experimental design can be improved.

Story Complexity and Presentation We suspect that one important source of inaccuracy and variance was the convoluted nature of the stories. In order to control all the factors identified in Section 4, the stories often switched between characters, locations, and times in ways that seemed unnatural. We considered this an acceptable risk given that we wished to test all five indices in the same experimental context, but it is certainly an area for improvement.

In three domains, we relied on the use of flashbacks to ensure a prompt shared only its time frame with the referent. It is not clear if a person's mental model treats a flashback as a return to a previous time fame (as to a previous location) or as the creation of a new time frame which happens to be in the past. We hope to explore this in future work.

Finally, the presentation of these stories as short, simple sentences limits their richness. Just as many EISM studies relied on visual feedback in addition to text, we would like to replicate this study in a rich audiovisual interactive story environment which will may affect the salience of events in the audience's memory.

Probing via Flavor Text Another source of inaccuracy may be the link, or lack thereof, between an event and its flavor text. We used flavor text as a means of indirectly probing a participant's memory to ensure that the probe text itself did not affect their mental model of the situation. It is possible that participants remember the flavor text independently of the referent to which it was attached. Addressing this narrative Heisenberg uncertainty principle is one of the critical aspects of the experimental design that needs to be addressed: How can we accurately probe for the salience of a past event without modifying the participant's mental model in the process?

Independence from Intervening Events We suspect that the most prominent flaw in the Pairwise Event Salience Hypothesis is that it ignores the events which occur between the referent and prompt. Initial work by Winer et al. (2015) demonstrates that the order in which events are presented in a plan-based story affects how their causal relationships are perceived. This may explain why the null hypothesis for causality was not rejected, and more generally it suggests the order and content of intervening events cannot be ignored.

The stories developed and the data collected for this study are not just valuable for exploring this hypothesis; they will also provide a testbed in which to examine more complex models. Cardona-Rivera et al. (2012) identified many nuances of EISM research that we have ignored in this evaluation. For example, indices may differ in how they affect salience, and they may interact when grouped together. This is demonstrated by the fact that participants reacted slower to *intentionality* than *protagonist* despite the fact that *intentionality* subsumes *protagonist* by definition. We represent events as either sharing five indices or not (a binary vector of length 5), but a more suitable representation might be a vector of real numbers which decay as the narrative progresses and other indices are activated by intervening events.

7 Conclusion

Indexter is a plan-based computational model of narrative discourse which operationalizes cognitive science research on how people remember past narrative events. This paper presented the first in a planned series of evaluations that will test increasingly nuanced models of how these indices can be used to predict salience.

The Pairwise Event Salience Hypothesis states that a past event is more salient when it shares one or more indices with the most recently narrated event. We found support for this simple model when considering the indices individually. However, the low accuracy and high variance in responses motives the development of a more nuanced model. Specifically, we intend to consider how intervening events and episodic memory decay affect the salience link between a pair of events.

If successful, the development of an accurate procedure for using *Indexter* to predict salience will have a significant effect on the interactive narrative community, especially for plan-based narrative generation techniques into which the model can readily be integrated. The ability to control salience will advance the capabilities of narrative generation systems to control comprehension and create suspense, surprise, misdirection, and other discourse phenomena.

References

Bae, B.-C., and Young, R. M. 2014. A computational model of narrative generation for surprise arousal. *IEEE Transactions on Computational Intelligence and Artificial Intelligence in Games* 6(2):131–143.

Bal, M. 1997. *Narratology: introduction to the theory of narrative*. University of Toronto Press.

Benjamini, Y., and Hochberg, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)* 289–300.

Brewer, W. F., and Dupree, D. A. 1983. Use of plan schemata in the recall and recognition of goal-directed actions. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 9(1):117.

Cardona-Rivera, R. E.; Cassell, B. A.; Ware, S. G.; and Young, R. M. 2012. Indexter: a computational model of the event-indexing situation model for characterizing narratives. In *Proceedings of the 3rd Workshop on Computational Models of Narrative*, 34–43.

Cardona-Rivera, R. E.; Robertson, J.; Ware, S. G.; Harrison, B.; Roberts, D. L.; and Young, R. M. 2014. Foreseeing meaningful choices. In *Proceedings of the 10th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, 9–15.

Cheong, Y.-G., and Young, R. M. 2008. Narrative generation for suspense: modeling and evaluation. In *1st International Conference on Interactive Digital Storytelling*, 144–155.

Fikes, R. E., and Nilsson, N. J. 1972. STRIPS: a new approach to the application of theorem proving to problem solving. *Artificial Intelligence* 2(3):189–208.

Gillund, G., and Shiffrin, R. M. 1984. A retrieval model for both recognition and recall. *Psychological Review* 91(1):1.

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

Macmillan, N. A., and Creelman, C. D. 2004. *Detection Theory: a user's guide*. Taylor & Francis.

Magliano, J. P.; Miller, J.; and Zwaan, R. A. 2001. Indexing space and time in film understanding. *Applied Cognitive Psychology* 15(5):533–545.

Morrow, D. G.; Bower, G. H.; and Greenspan, S. L. 1989. Updating situation models during narrative comprehension. *Journal of Memory and Language* 28(3):292–312.

Reimers, S., and Stewart, N. 2014. Presentation and response timing accuracy in Adobe Flash and HTML5/JavaScript web experiments. *Behavior research methods* 47(2):309–327.

Riedl, M. O., and Young, R. M. 2010. Narrative planning: balancing plot and character. *Journal of Artificial Intelligence Research* 39(1):217–268.

Trabasso, T., and Sperry, L. L. 1985. Causal relatedness and importance of story events. *Journal of Memory and language* 24(5):595–611.

Trabasso, T., and Van Den Broek, P. 1985. Causal thinking and the representation of narrative events. *Journal of memory and language* 24(5):612–630.

Winer, D. R.; Amos-Binks, A. A.; Barot, C.; and Young, R. M. 2015. Good timing for computational models of narrative discourse. In *Proceedings of the 6th International Workshop on Computational Models of Narrative*, 152–156.

Yonelinas, A. P., and Jacoby, L. L. 1994. Dissociations of processes in recognition memory: effects of interference and of response speed. *Canadian Journal of Experimental Psychology* 48(4):516.

Young, R. M.; Ware, S. G.; Cassell, B. A.; and Robertson, J. 2013. Plans and planning in narrative generation: a review of plan-based approaches to the generation of story, discourse and interactivity in narratives. *Sprache und Datenverarbeitung, Special Issue on Formal and Computational Models of Narrative* 37(1-2):41–64.

Zacks, J. M.; Speer, N. K.; and Reynolds, J. R. 2009. Segmentation in reading and film comprehension. *Journal of Experimental Psychology: General* 138(2):307.

Zacks, J. M.; Tversky, B.; and Iyer, G. 2001. Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General* 130(1):29.

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