A Cognitive-Based Model of Flashbacks for Computational Narratives

Hui-Yin Wu, R. Michael Young and Marc Christie
IRISA/INRIA Rennes Bretagne Atlantique and North Carolina State University
hui-yin.wu@inria.fr, rmyoung@ncsu.edu, marc.christie@irisa.fr

Abstract
The flashback is a well-known storytelling device used to invoke surprise, suspense, or fill in missing details in a story. Film literature provides a deeper and more complex grounding of flashbacks by explaining their role to stimulate the viewer’s memory in order to guide and change viewer comprehension. Yet, in adapting flashback mechanisms to AI storytelling systems, existing approaches have not fully modelled the roles of a flashback event on the viewer’s comprehension and memory.

To expand the scope of AI generated stories, we propose a formal definition of flashbacks based on the identification of four different impacts on the viewer's beliefs. We then establish a cognitive model that can predict how viewers would perceive a flashback event. We finally design a user-evaluation to demonstrate that our model correctly predicts the effects of different flashbacks. This opens great opportunities for creating compelling and temporally complex interactive narratives grounded on cognitive models.

Introduction
Genette (1980) says flashback was a feature of memory before it was of narrative; Bordwell (1985) says it’s there to remind us; Turim (1989) further characterizes flashbacks with the role of enlightening, haunting, surprising, and changing our beliefs towards story events. Having much to do with memory, inserting events as flashbacks throughout the story can guide the viewer to establish causal relations between events that are chronologically distant, and to remember (or forget) them at a specific pace.

Flashbacks have a role of guiding the viewer’s comprehension of events. Current AI storytelling systems can generate flashbacks that invoke surprise and suspense (Bae and Young 2008), and provide focalization (Hoek, Theune, and Linssen 2014)(Montfort 2011). These works address flashbacks as an event sequencing problem: if every event appears once, in what order should we place them? Flashbacks however have a much wider range of cognitive effects on the viewer and have not been fully addressed in the storytelling literature. The film theorist Maureen Turim (1989) finds that flashbacks have the cognitive memory function of guiding the viewer’s perception: establishing new beliefs, changing or reinforcing them to invoke an emotional response, and reminding through repetition. As an inserted event, flashbacks can help the viewer establish causal relations to close and salient events. Yet, how does an AI system select an event that provides the author’s intended effect to the audience? How do we know that it does indeed guide the viewer’s understanding of the story? These aspects have been insufficiently explored in existing storytelling systems.

Thus, the purpose of this work is to answer three fundamental questions concerning flashbacks in AI storytelling: (A) What is a flashback? (B) What is the effect of a flashback on the viewer’s memory? (C) How can we evaluate the range of choices an author can make when inserting an event as a flashback? In this paper we design an algorithm which provides a more general formulation of flashbacks than done in previous work, evaluate the effect of flashbacks with a wider range of cognitive functions, and validate the generated output through a user evaluation.

Related Work
In film theory, the flashback is broadly defined as a technique in which an earlier event is inserted into an otherwise chronological sequence (Genette 1980)(Turim 1989)(Bordwell 1985). Turim (1989) analyses the use of flashbacks throughout film history, and finds that the flashback is a visual device, and has much to do with memory.

The temporal rearrangement of events has been a study of AI storytelling systems, both in cognitive and generative aspects. Lönnker(2005) and Wei (2010) portray flashbacks as embedded narratives focusing on game story structure with applications to natural language and game flow design. (Winer et al. 2015) proposes a framework and an analysis of structural properties of discourse to allow reasoning on timing. However, this does not address viewer comprehension, memory or flashbacks, and was not fully evaluated. A number of works target flashback generation. Montfort (2007) proposes an algorithm to reorder events in text stories while respecting correct grammatical tense of the output order of events, which they later use to generate flashbacks focalised from a specific character’s viewpoint (Montfort 2011). Flashbacks for character focalisation is also addressed by Hoek et al.(2014) to create backstories of actors. Bae and Young (2008) develop a methodology to generate
flashback and foreshadowing to invoke surprise. So far, these approaches focus on finding a sequence of events that satisfies inter-event causal constraints.

Our approach sets out from a cognitive perspective. Like in previous work on suspense in stories (O’Neill and Riedl 2014), we use a cognitive model for calculating memory retention and salience of events (and associated beliefs). However, our approach is targeted not towards the generation of flashbacks, but the exploration of authorial creative choices, identifying a broader and more general range of cognitive effects of inserted flashback events: establishing, reinforcing, and changing viewer beliefs.

**Overview**

The purpose of this work is to establish a cognitive model of flashbacks for computational narratives, and to evaluate the different types of changes on the viewer’s beliefs flashbacks result in. We design an algorithm that takes as input the story (a collection of events) and the belief to be evaluated at a target discourse time (e.g. “Jerry knows movie A” after 30 minutes into the film). The algorithm (a) searches for events that alter this belief, and for all events evaluates each possible insertion point considering the memory retention of the belief at the target discourse time, and (b) determines which type of change on the belief the flashback corresponds to.

The rest of our contribution is structured as follows: we first present our story representation on which to build our flashback mechanism, our model of viewer memory, and our definition of beliefs. We then follow with the algorithm for exploring possible flashback constructions and their varying impacts on the viewer’s beliefs. The output of the algorithm and user evaluations are presented in the results section. Finally, we discuss the limitations and future work.

**Story, Memory, Beliefs and Goals**

We propose the following representations and definitions on which we build our contribution.

**Story and Temporal Representation**

Two timelines are distinguished: story time and discourse time. Story time represents what actually happened in the world the actors live in, their time, their events. Discourse time represents what the viewer perceives, in the viewer’s time, the sequence in which story events are shown in a movie.

We view the story as a collection of pre-authored events represented by propositions (e.g. $P$=“Jerry goes to movie A.”). Due to the importance of temporal relations between events in flashbacks, we rely on the format of temporal representation proposed in (Eger, Barot, and Young 2015), which annotates a start and end timestamp in story time during which the event spans (e.g. “Jerry goes to movie A.”; start:18h00 16 November 2015; end:20h30 16 November 2015). The notation $P_{st}$ refers to a proposition $P$ which occurs during the story time interval $st$. The temporal information helps to establish relations between events in story time by reasoning on Allen interval relations (Allen 1984), and to ensure the flashback event being prior in story time.

**Beliefs**

When an event is shown to a viewer, the event invokes certain beliefs in the viewer about the story world for a certain story time. For example, the event $P_{st}$ = “Jerry goes to movie A.” at time interval $st$ could make viewers believe “Jerry knows movie A.” is positive, starting from the end time of $st$ (after Jerry viewed the film). Our story event representation associates to each event a number of beliefs.

**Definition 1 (Belief).** A belief $b(P_{st})$ is the viewer’s perception of a proposition $P$ occurring during $st$. A belief is composed of three properties:

- **Value** of the belief of the viewer is either positive (“believe that”), negative (“believe not that”), or ignorant.

- **Proposition** of the belief, such as “Jerry knows movie A.”

- **Story Time** interval of the belief (the story time for which the viewer believes the proposition, see Figure 1).

The viewer has a set of beliefs that are changed and updated as the story progresses. The memory state of the viewer at any given discourse time is therefore defined as:

**Definition 2 (Memory state).** The memory state of a viewer $v$ at time $t$ is represented as the set of beliefs $B_{v,t}$ of the viewer $v$ at time $t$. A belief $b_{v}(S_{st})$ amongst beliefs $B_{v,dt_{n}}$ for a viewer $v$ at discourse time $dt_{n}$ corresponds to the belief that a proposition $P_{st}$ that occurred at story time $st$ is positive or negative at time $dt_{n}$ in the viewer’s memory

$$\quad (P_{st} \lor \neg P_{st}) \in B_{v,dt_{n}}$$

(1)

If the proposition $P_{st}$ is ignorant (as are all beliefs at the beginning of the story), $P_{st}$, neither positive or negative, is not in $B_{v,dt_{n}}$:

$$\quad (P_{st} \lor \neg P_{st}) \notin B_{v,dt_{n}}$$

(2)

Initially, we consider the status of all beliefs in the viewer to be ignorant. As the story unfolds, beliefs can be established (from ignorant to positive or negative), changed (from positive to negative, or conversely), or forgotten (from positive or negative to ignorant). Beliefs are effective or ignorant at specific story times based on their timestamp (e.g. “Jerry knows movie A.” is effective for the rest of the story after “Jerry goes to the movie A.”).

**Memory retention and salience**

Not only are beliefs associated with story events (propositions) in story time, but they can also be forgotten (i.e. become ignorant again) as discourse time increases. Events and their associated beliefs in the viewer’s memory are modelled to calculate this decay.

According to research on the temporal theory of memory decay, it is found that working memory has an exponential decay in proportion to the strength of that memory. The strength is defined as its salience—the state and quality that makes an item stand out from its neighbours (Ebbinghaus 1885)(Buhusi and Meck 2006)(Averell and Heathcote 2011). The probability that a person will recall an event a certain amount of time after it is introduced is termed memory retention. The function for memory retention ($R$) in relation to time ($T$) and memory strength/saliency ($S$), also...
referred to as forgetting curve, is defined by Ebbinghaus (1885):

\[ R = e^{-\frac{T}{T_0}} \]  

(3)

To estimate the saliency, we rely on the Event Indexing Model (McNamara and Magliano 2009), an empirically verified model of how viewers interpret story events along the indices of time, space, character, intention, and causality between events in the story. This representation models the saliency of events in film narrative, and was further developed into a computational EISM model by Cardona-rivera et al. (2012), using the above indices in order to calculate the saliency of an event in a sequence of events. In our story representation, we propose to add EISM links between two events \( E_i \) and \( E_n \) in the story, with indices of time \( (t) \), location \( (l) \), character \( (ch) \), intention \( (in) \), and causality \( (ca) \), which we assume are authored or can be generated. We calculate the saliency between two events \( E_i \) and \( E_n \) as the sum of these indices multiplied by evenly distributed weights as proposed in (Cardona-rivera et al. 2012).

\[ S(E_i, E_n) = 0.2 \times t_{E_i, E_n} + 0.2 \times l_{E_i, E_n} + 0.2 \times ch_{E_i, E_n} + 0.2 \times in_{E_i, E_n} + 0.2 \times ca_{E_i, E_n} \]  

(4)

For example, the indice \( t_{E_i, E_n} \) represents the degree to which \( E_i \) and \( E_n \) are temporally connected (defined between 0=not connected and 1=connected). At a given discourse time of the story, knowing the saliency \( S \) between events and the discourse time \( T \), we can then compute the memory retention value \( R \) that represents the decaying strength of a belief to a viewer. The model is simple, but has a basis in cognitive theory, and fulfills our purpose to (1) continuously evaluate the viewer’s memory state along the discourse, and (2) guide viewer comprehension and interpretation of causality between events.

Up to this point, we have differentiated between story time (time in story world) and discourse time (time of the telling of the story), discussed how our story representation models events, how events can have associated beliefs, and how the saliency of these events and associated beliefs are calculated using the EISM model. The way these events and beliefs are modelled in our system is summarised in Figure 1.

**Algorithm: Flashback and Impact on Beliefs**

Given our story representation, we now explore the issue of inserting a story event as a flashback with two questions: (1) How relevant is an inserted event to its neighbouring events? (2) What is the effect of inserting the event on the viewer’s beliefs? The first question of relevance to neighbouring events concerns the salience of the event to the viewer. If the event is more salient, it would be easier for viewers to establish cause-effect relation between the inserted event, and other neighbouring or close events.

**Insertion points**

Events can be inserted as a flashback at insertion points between two events. Possible insertion points can be numerous. We rank the score of the insertion point by calculating the salience between the inserted event and the neighbouring events. The higher the score, the better the quality of the insertion point. The score for an insertion point \( i_{E_1, E_2} \) neighboured by events \( E_r \) and \( E_s \) is calculated using the salience value in Equation (3):

\[ \text{Definition 3. InsertPointScore}(E_1, i_{E_1, E_2}, E_r, E_s) = 0.5 \times (S(E_1, E_r) + S(E_1, E_s)) \]

**Cognitive Model for Flashbacks**

The relation between the inserted flashback event with the rest of the story can result in varying interpretations by the viewer. From (Turim 1989), we simplified the types of flashback events in terms of the ways flashbacks influence the viewer’s beliefs. Four types of changes on the viewer’s beliefs are identified, namely:

- **establishing** a new belief previously ignorant to the viewer;
- **reinforcing** of an event/belief to refresh a belief that is no longer salient;
- **changing** the value of a salient belief from positive to negative (or vice versa);
- **salient** : the inserted flashback does not result in a change of memory state.

For simplicity, we refer to a flashback by the type change it results in (e.g. establishing flashback, reinforcing flashback...etc.). We provide examples of each kind of flashback from *Lord of the Rings: The Fellowship of the Ring*.

**Salient** means that inserting the event at the specified insertion point adds a belief that already exists with higher than 50% retention. A salient flashback is easily defined as:

**Definition 4 (Salient flashback).** If \( E_1 \), inserted at \( dt_m \), has a retention of \( E_1, R(t = dt_m - dt_m) \geq 0.5 \), the belief is still in the viewer’s memory, and the flashback has no effect on the viewer.

Throughout the *Fellowship*, frequent flashbacks of Sauron’s eye are more for dramatic effect than for the need to remind the viewer of its link to the ring.

![Figure 2: Establishing flashback: if Event 1 is inserted as a flashback between Event 4 and Event 5, this establishes a new belief \( \text{bel} \) in the viewer's memory at discourse time.](image-url)
Establishing means that the viewer establishes a belief that some proposition is true where the viewer was previously ignorant of that proposition, thus increasing both the memory retention and the saliency of the event’s associated beliefs at the target discourse time.

Definition 5 (Establishing flashback). An establishing flashback is where some proposition $P$ about story time interval $st$ is inserted at discourse time $dt_n$, and that $P_{st}$ was not in the set of beliefs $B_v$ of the viewer $v$ at any discourse time $k < n$.

$$\forall k < n : P_{st} \notin B_v,dt_n \land (P_{st} \in B_v,dt_n) \quad (5)$$

The visualization of the establishing flashback can be found in Figure 2. In the Fellowship, Elrond narrates Isildur’s fall in a flashback, establishing the belief that Elrond distrusts men.

Changing indicates that a belief already held by the viewer towards some proposition is negated. In the memory model, the belief itself changes value (negated), but the memory retention and saliency is unchanged.

Definition 6 (Changing flashback). A changing flashback is when some proposition $P$ about story time interval $st$ is inserted at discourse time $dt_n$, and $P_{st}$ replaces the salient belief $\neg P_{st}$ in the set of the viewer beliefs $B_v$ at $dt_n$.

$$(\neg P_{st} \in B_v,dt_{n-1}) \land (P_{st} \in B_v,dt_n) \quad (6)$$

The visualization of the changing flashback can be seen in Figure 3. In the Fellowship, when Gandalf narrates the danger of the One Ring, Frodo verbalises the belief that Sauron does not know where the ring is. A flashback of Gollum being tortured by Sauron soon negates this belief.

Reinforcing is where the viewer believes some proposition that was shown previously but forgotten, which means that the memory retention of an event is renewed.

Definition 7 (Reinforcing flashback). A reinforcing flashback is when some proposition $P$ that is about story time interval $st$ is inserted at discourse time $dt_n$, and that belief $P_{st}$

is in the set of beliefs $B_v$ of the viewer at some discourse time $k \leq n$ but not in $B_v$ at discourse time $dt_n$.

$$(P_{st} \notin B_v,dt_{n-1}) \land (\exists k < n : P_{st} \in B_v,dt_k) \land (P_{st} \in B_v,dt_n) \quad (7)$$

A visualization of a reinforcing flashback is similar to the establishing flashback, except the inserted event would repeat a non-salient event (such as Event 3 in Figure 2). In the Fellowship, Isildur cutting the ring off Sauron’s finger appears twice in the film: once in the beginning of the film, and again in the middle of the film just before Elrond provides new information on Isildur’s fall.

Algorithm

We design Algorithm 1 from our formal definition of flashbacks, involving three main steps: find events that are relevant to a given belief at a given discourse time (e.g. events related to the belief “Jerry knows movie A” that should be positive at discourse time $dt$), rank salient insertion points for these related events, and calculate the type of change on the viewer’s beliefs.
The algorithm takes as input a story $mathcal{S}$ (a collection of events), the belief we would like to observe, and a target discourse time at which the belief should occur. First, the algorithm searches for all candidate events that possibly alter the belief by first testing whether the candidate event impacts the belief $b$, and if the story time of the belief is during (in terms of Allen time intervals) the story time of the belief invoked by the candidate event. It then evaluates the insertion point score of each candidate event/insertion point pair. The retention $R$ of this event is computed at time $T = dt - E_{dt}$ representing the amount of time that has passed between showing the event $E$ and the target discourse time $dt$. From Definitions 5~8 (Algorithm 1. Line 12), we deduce the type of change on the viewer’s memory the flashback has (function $flashbackType$). Finally, we return a ranked list event-insertion point pairs (by insertion point score), and the flashback type.

**Results**

To evaluate the output of our algorithm, we annotated events from the film synopses of Constant Gardener (CG), Lord of the Rings (LotR), and one hand-crafted story “The Interview”. We chose the same scenario in LotR as in (Eger, Barot, and Young 2015) to demonstrate how our algorithm evaluates the effect of the inserted flashback event on the viewer. Temporal information was collected from the original chronology provided by the author, J.R.R. Tolkien, but names were changed to reduce association with the original material.

Figure 4 shows a sample variation of “The Interview” computed by our algorithm with viewer beliefs $\neg P_{S13-15} \in B_{v, dt:10}$ and $P_{S13-15} \in B_{v, dt:15}$ where $S13-15$ means the story time of scenes 13-15, $dt:15$ means discourse time at Scene 15, $P$="Jerry is confident for his interview.”

For each story, we presented participants with three variations: the one recommended with highest event-insertion point score from Algorithm 1, the film version, and the linear version. We tested two hypotheses. **H1:** do our algorithm correctly evaluate how changing the order in which events are presented reflects the viewer’s interpretation? **H2:** can the cognitive model correctly predict the type of flashback the viewer perceives?

We recruited 41 participants: 22 males and 19 females, ages ranging 19-37. Three variations of each story—recommended, film, and linear—were presented to participants in random order. For our hand-created story, we presented two generated versions with different flashbacks. Participants were given a list of 1 sentence scene descriptions as in Figure 4. Since we refer to the flashback technique in terms of film, participants were asked to imagine to be the viewer at a movie. All participants read at least one version of each story, and we ensured no two variations of the same story appeared consecutively. In the end, each variation was evaluated by 24-30 subjects.

For **H1**, we asked participants to identify the scene that they felt was the salient cause of a specific Scene X. In each story variation, multiple Scene Xs were demanded. By asking this question, our purpose is to see if participants would identify the same scenes as salient causes, when they appeared as a flashback, or when they appeared chronologically. Figure 5 shows the relation between the memory retention (normalized for each story variation such that the sum of all scenes is 1.0) for the scenes that were identified by participants as salient, and the number of participants that identified the scene as a salient cause. Each dot represents a scene in the story, on the $x$-axis, its memory retention percentage (from Definition 2), and on the $y$-axis, the percentage of participants that selected the scene as being the salient cause. In each graph for the stories, the flashback events are coloured in red.

On the whole, our findings are as follows: First, there is a positive correlation between the memory retention and the participant responses, the highest $r$-square value up to 0.65 in our hand-crafted story, and 0.32 over the whole dataset. For an experiment involving human participants, an $R$-squared value lower than 50% is expected, due to humans being harder to predict (Frost 2013). Thus our data is sufficient to show that the calculation of memory retention percentages of events does reflect whether participants found the events salient or non-salient at different points of the story. Second, we found that the difference between inserting an event as a flashback and presenting it in chronological order changes the viewer’s interpretation of the overall plot line. In the top part of Figure 5, the green circles we have annotated indi-
Figure 5: Each dot represents a scene selected as a salient cause to a target scene by the participant, and the red dots in the top graph represent those that are flashbacks. Relation between the normalized memory retention for a scene (normalized for each story variation such that the sum of all scenes is 1.0), and percentage of participants who chose a scene as the salient cause are shown.

cate the same scene—Scene#2 in Figure 4—that was chosen by participants as the salient cause of the same target scene in the three variations of our hand-crafted story. It shows that when Scene#2 is inserted as a flashback closer to our target event, a higher percentage of participants identified it as a salient cause. We found the same results with other target events, and for the two other stories, Lord of the Rings and Constant Gardener. This shows our algorithm can correctly evaluate the change in the viewer’s belief when the order of events is changed in the story, and H1 is validated.

For H2, three statements corresponding to the Establishing, Reinforcing/Salient, and Changing effects on viewer beliefs were presented to the participants. They could select any number of statements that they thought appropriate to describe the flashback. Figure 6 shows the distribution of participant responses for each flashback type to the type identified by the algorithm (as indicated by the label). We found that the participant responses strongly correspond to the flashback type identified by our algorithm: Especially, all establishing and reinforcing flashbacks show the majority of participant responses fall in the correct category, and 2 of 3 changing flashbacks were also significant. Thus, H2 is validated.

Discussion and Conclusions

We have presented a cognitive model to explore the flashback choices an author can make (inserting an event as a flashback at a certain point) to achieve a desired belief, and how these choices affect the viewer’s perception through different flashbacks (establish, change, reinforce, or salient). While our approach is computationally simple, we show that our cognitive-based modelling of flashbacks correctly evaluates and guides viewer’s interpretations.

There are a number of limitations to address. The model of saliency on was empirically verified, yet memory in storytelling remains not well-understood, and thus, it is still far from predicting exactly how a viewer would interpret flashbacks. A step in a direction is the work of (Kives, Ware, and Baker 2015). Secondly, our evaluation is designed to understand how sequence and proximity of events alone affects interpretation. To limit control factors, our evaluation was in text form which cannot simulate temporal features of flashbacks in film such as duration.
References


Eger, M.; Barot, C.; and Young, R. M. 2015. Impulse: a formal characterization of story. In *Sixth International Workshop on Computational Models of Narrative (CMN 2015)*.


