Manipulating Narrative Salience in Interactive Stories Using Indexter's Pairwise Event Salience Hypothesis

Rachelyn Farrell, Stephen G. Ware, *Member, IEEE*, and Lewis J. Baker

***Abstract*—The salience of a narrative event is defined as the ease with which an audience member can recall that past event. This paper describes a series of experiments investigating the use of salience as a predictor of player behavior in interactive narrative scenarios. We utilize Indexter, a plan-based model of narrative for reasoning about salience. Indexter defines a mapping of five event indices identified by cognitive science research onto narrative planning event structures. The indices—protagonist, time, space, causality, and intentionality—correspond to the "who, when, where, how, and why" of a narrative event, and represent dimensions by which events can be linked in short-term memory. We first evaluate Indexter's claim that it can effectively model the salience of past events in a player's mind. Next, we demonstrate that salience can be used to predict players' choices for endings in an interactive story, and finally, we demonstrate that the same technique can be applied to influence players to choose certain endings.**

I. Introduction

Interactive narratives allow players to engage with the narrative environment in ways that affect the resulting story. To ensure that the narrative stays on track despite unexpected input from the user, many interactive narrative systems contain an experience manager component that monitors the player's experience and decides which content to display at each point [1]. Typically this is done by controlling nonplayer characters and choosing other world-level actions (e.g., locking doors, or placing things where the player will find them).

Experience managers often utilize player models—structures that contain information or assumptions about the player and can be used to predict the player's future actions [2]–[5]. Player models can be generic; e.g., a player who explores the world randomly, or a player who exhausts all possibilities in a single location before moving on. They can also be based on theoretical player types, e.g., a fighter or a pacifist. Some systems use dynamic player models that are updated throughout the scenario as more information about the player's behavior becomes available [4], [5]. Models can also be informed by the user's previous playthroughs or by data from other users.

One strategy is to maintain a detailed representation of the player's knowledge or understanding of the story based on the story's discourse. Knowledge representation is especially useful in educational scenarios or in certain genres, such as detective stories and mysteries, where information is key. By modeling what the player knows and does not know, the system can explicitly choose actions that reveal or reinforce important details to the player. User knowledge models can not only represent what players currently believe, but also what they expect from the future and even how they imagine hypothetical situations would turn out [6]. Still, knowledge structures alone fall short of representing what players *desire* from the future or what preferences or intentions they have.

In this work, we investigate Indexter [7], a computational model of narrative event salience (or prominence in memory) as a means of predicting and influencing player behavior. Based on the story's events and discourse, we can model which past events are more salient in the player's memory at a given time. To demonstrate that this can be useful in player modeling, we conducted a study in which we predicted readers' choices for the final event in the story based on the salience of past events. In a follow-up study, we showed that we can also influence players' choices for final events by manipulating which past events would be salient at the time.

Indexter is based on the event-indexing situation model (EISM) [8], which states that narrative events are stored and retrieved in short-term memory along at least five indices: *protagonist, time, space, causality,* and *intentionality* (in other words: *who*, *when*, *where*, *how*, and *why*). Indexter proposes a mapping of these indices onto narrative planning events (discussed in detail in Section III). This mapping allows plan-based narrative systems to calculate the salience of any past event as a function of the indices it shares with the current event. For example, if a scene is currently taking place in the Tavern, then a previous scene which also took place in the Tavern should be more salient than others because it shares the *space* index with the current event.

Although more nuanced models of narrative comprehension have been proposed [9], we use the simplest interpretation of Indexter (detailed in Section IV-A). We do this in part because more research is needed to evaluate other options. Furthermore, we assume that any effect we see with the simplest model can

Manuscript received January 3, 2018; revised January 8, 2019; accepted February 13, 2019. Date of publication March 15, 2019; date of current version March 17, 2020. This work was supported by the National Science Foundation (NSF) under Award IIS-1464127. *(Corresponding author: Rachelyn Farrell.)*

R. Farrell and S. G. Ware are with the Department of Computer Science, University of New Orleans, New Orleans, LA 70148 USA (e-mail: rfarrell@uno.edu; sgware@uno.edu).

L. J. Baker is with the Department of Psychology and Human Development, Vanderbilt University, Nashville, TN 37203-5721 USA (e-mail: lewis.j.baker@vanderbilt.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TG.2019.2905230

be strengthened by using a more complex one. We refer to this as the *pairwise event salience hypothesis*:

> A past event is more salient when it shares one or more indices with the current event than when it shares no indices.

As a starting point for how salience of past events relates to players' preferences for future events, we used the following hypothesis:

> Players are more likely to choose a future event which makes the past (overall) more salient.

This is inspired by the idea that players want the story to feel consistent and unified, and that endings can remind players of previous events as a way of bringing closure or "wrapping up" the narrative.

This paper describes a series of studies that evaluates Indexter as a measure of salience and suggests a means of relating the salience of past events to preferences for future events. Specifically, we make the following claims.

1) We can effectively approximate the salience of past story events in the audience's memory using Indexter.
2) We can predict players' choices for ending events based on which past events are more salient.
3) We can influence players to choose certain endings over others by manipulating event salience.

In this paper we describe three previously published studies ([10]–[12]) and present their results, present additional results from three previously unpublished variations of the third study (Sections IV-A3–IV-C3), and provide a unified discussion of the studies and their findings.

II. Related Work

Much of the research in the interactive narrative landscape focuses either on character believability [13]–[15] or on adapting story structure and content in response to user input [1], [16], [17]. Researchers have identified a need for a central *experience manager* (or drama manager) that dynamically selects content to display in response to user interaction. Experience managers have used a variety of techniques [18]–[22] and often make use of player models to predict which future plot points the player is likely to choose. Our work relates specifically to research toward building computational models of users' knowledge, expectations, and preferences; and in particular, those based on psychological models of narrative comprehension through discourse.

Nelson *et al.* implemented a declarative optimization-based drama manager (DODM) that relies on predictions of likely player behavior [23]. They used two simple player models—that of a player exploring the world randomly, and that of a player who explores randomly, but also listens to suggestions given by the narrative. Their evaluation of the DODM showed encouraging results when reinforcement learning was used to precompute the decision policy offline. Other player modeling systems, such as Thue *et al.*'s PaSSAGE [5] classify players into predefined behavioral categories based on their actions. This approach showed promising results, although it relies on accurate classification of players. This can be difficult in practice because players may initially exhibit exploratory behavior—taking actions essentially randomly—rather than acting in accordance with the player type that normally best suits them.

There are comparatively few works that focus specifically on influencing user behavior [24]. Roberts and Isbell were able to influence players using player models and concepts from social psychology, discourse analysis, and natural language generation [25]. El Nasr *et al.* influenced players in game environments by using lighting and other techniques to draw their attention to important elements in the game [26]. Our approach models users' preferences based on the salience of related events in their memory. We believe this technique can be combined with other player modeling strategies to produce a robust model that more accurately predicts the player's preferences and future actions.

Modeling player knowledge has been studied primarily in the context of education and training systems or mystery genres, where the user's understanding of the narrative and its content is central. Rowe and Lester presented a method for modeling users' knowledge with dynamic Bayesian networks (DBN) [27]. Their method was able to predict users' accuracy on a postexperiment test more effectively than a random baseline. This is a promising technique, but the construction of the dynamic Bayesian networks requires either machine learning from data or extensive hand authoring about the relevant knowledge components.

An alternative approach is to use QUEST, a psychological model of question answering [28], to represent the user's knowledge of story events. Researchers have developed a mapping of QUEST knowledge structures onto narrative plans which has been used to validate various plan-based models of narrative [6], [29], [30]. Additionally, players' expectations about future story events have been studied in the context of narrative affordance—the idea that certain courses of action are suggested to be available based on the narrative and its genre [31]. In our work, we focus on players' short-term memory about past events and how this affects their preferences for the future, operationalized as which actions they would choose from a set of known options.

Our approach to linking salience with future choices is related to the improvisational theater technique of reincorporation [32], in which improv actors make previous events become narratively necessary by introducing new events that logically depend upon them. An evaluation of the use of reincorporation in the interactive drama system Marlinspike [33] showed that although reincorporation improved story unity and incorporated more of the player's actions into the main story thread, it did not have a significant effect on players' experience. Our salience-based technique differs from reincorporation in the way that we "reference" past events; rather than incorporating them causally into the current event, we are simply adjusting the details of the current event (the who, when, where, how, and why) such that the player is reminded of the previous event.

Our method is based on Indexter [7], a plan-based model of narrative discourse for reasoning about salience. Plan-based narrative systems [34] are ideal for reasoning about story structure, particularly in the context of interactive stories with large branching factors where hand authoring all the content becomes infeasible. Plan-based models can incorporate character intentionality [35], conflict [36], and individual character beliefs [37].

Plan-based models have also been used to analyze specific discourse elements, including suspense [38], surprise [39], and cinematic composition [40]. An overview of narrative planning is given in Section III.

Indexter has also been used to predict how much agency—the ability to have a meaningful impact on the narrative—players feel when making choices [41]. The results of that study are particularly relevant to the evaluation of our third claim and will be described in Section IV-C.

III. Indexter Model

Indexter defines a mapping of the five event-indexing situation model (EISM) indices—*protagonist*, *time*, *space*, *causality*, and *intentionality*—onto narrative planning structures, such that a pair of events may share up to five of these dimensions with each other. This section first gives a brief overview of narrative planning concepts, then describes Indexter's mapping in detail.

A planner is an algorithm that attempts to solve this 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. Our model is STRIPS based, meaning the kinds of events that can occur are represented by abstract, parameterized operators [42]. 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 plan, or sequence of events, that will achieve the goal when executed from the initial state.

Narrative planning is an application of planning for generating stories. The problem domain encodes the available characters, objects, locations, etc. The initial state describes how the story begins, and the goal describes what is required for the story to end. The operators define any type of event or action that may occur in the story. For example, if a character may steal an item from another character, we might define an operator like this:

steal(?thief, ?owner, ?item, ?location)

Each parameter is a free variable which can be bound to a constant corresponding to some specific thing defined in the story domain. The preconditions for this event might be that both characters are currently at the same location, and that one character has the item, which could be represented as:

at(?thief, ?location) &

at(?owner, ?location) &

has(?owner, ?item).

This event would have the effect that the thief now has the item, and the owner no longer does:

has(?thief, ?item) &

¬ has(?owner, ?item).

The remainder of this section describes, for each index, how Indexter determines whether two events share that dimension.

Space and Time: Two of the indices, space and time, are mapped directly to an event's parameters. Indexter requires that all domain operators contain a ?place parameter representing the location that the event occurs, as well as a ?time parameter representing the time frame at which it occurs. Given this constraint, we can say that two grounded events share the *space* index if their ?place parameter is the same symbol, and the *time* index if their ?time parameter is the same symbol. This requires the domain author to determine the appropriate level of granularity at which to discretize locations and time frames.

Although cognitive science research has demonstrated that space and time can be hierarchically organized in memory [43], [44], we find that interesting results can still be obtained using this simpler representation of times and places as unique symbols. In our studies, we controlled for the ambiguity of spatial and temporal organization by explicitly communicating the times and places of events to the reader at the appropriate granularity.

It is worth distinguishing time in this context from recency. The *time* index for an event represents a relative time frame within the context of the story, such as "daytime" or "February." It is independent of how recently the event was narrated in the story's discourse.

Causality: For the causality index, we utilize causal links, which were originally developed for partial-order causal link planning [45]. A causal link signifies that there is a causal dependency between two events—specifically, an effect of the earlier event establishes a precondition for the later event. A causal link is formally defined as follows:

> 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 that s is the *causal parent* of t, and that an event' s *causal ancestors* are those events in the transitive closure of this relationship.

We can now say that, under the Indexter model, two events share the *causality* index when one is the causal ancestor of the other.

Protagonist and Intentionality: In addition to the author's goal, some narrative planners also reason about the goals of individual characters. Character goals need not be satisfied in the final state, but are used to justify the actions characters take. Riedl and Young's [35] model of character intentionality requires event operators to denote which, if any, of their parameters represent *consenting characters*, or characters who must "consent" to taking the action. In the *steal* example, the thief should be a consenting character, because he or she must have some motivation for stealing the item. While the owner of the item is also a character, and is also involved in the action, he or she is a passive participant and need not give consent in order for the event to be reasonable. Two events share the *protagonist* index if they have at least one consenting character in common.

Narrative planners using this intentionality framework keep track of commitment frames for each character. In general, a commitment frame is a series of causally linked actions that starts with the adoption of a character goal and ends with the satisfaction of that goal. A character may have many goals, and there may be many commitment frames for a single goal. To determine whether a character c consents to an action a, the planner searches c's commitment frames for occurrences of a.

If it finds one, this means that *c*has some goal *g* which can be used to explain *c*'s motivation to take action *a*.

The final index can now be defined as follows. Two events share the *intentionality* index if they appear in the same commitment frame—in other words, if both actions can be explained by the same character goal.

In summary, two events share the following:

1) the *space* index iff their ?place parameters are the same symbol;
2) the *time* index iff their ?time parameters are the same symbol;
3) the *causality* index iff the earlier event is a causal ancestor of the later event;
4) the *protagonist* index iff they have one or more consenting characters in common;
5) the *intentionality* index iff both events have a consenting character *c*, *c* has a goal *g*, and both events are causal ancestors of some event that has the effect *g*. In other words, both actions are taken by the same character in pursuit of the same goal.

IV. Studies

A. Measuring Past Event Salience

The original description of Indexter [7] proposed calculating the salience of a past event e_i based on the presence or absence of shared indices with the event currently being perceived. Each index that the current event e_n shares with the past event e_i is given a value of 1; other indices are given 0. Each index should also be assigned a weight coefficient corresponding to the strength of its individual contribution to salience, such that the total salience will be between 0 and 1. Then the salience of e_i can be calculated using the following:

$$\text{salience}(e_i, e_n) = w_1 t_{e_n} + w_2 s_{e_n} + w_3 p_{e_n} + w_4 c_{e_n} + w_5 i_{e_n}$$

where t_{e_n} is the *time* index of the current event e_n, and likewise s_{e_n}, p_{e_n}, c_{e_n}, and i_{e_n} are its *space*, *protagonist*, *causality,* and *intentionality* indices, respectively; and finally $w_1, \ldots, w_5$ are the weights of each index. The authors of the EISM do not specify the relative strengths of the indices, so until these values can be determined empirically (a clear direction for future work) we assume that the indices are equally weighted. That is, $w_j = 0.2$ for $j = 1 : 5$.

The purpose of our first study [10] was to validate this calculation by evaluating its claim that a past event becomes more salient when it shares indices with the current event. We formalized the *pairwise event salience hypothesis* as follows: 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 it.

For this evaluation we considered only events that shared exactly one or zero indices, leaving the "one or more" aspect of the hypothesis for future work. Because the binary comparison (any index versus no index) ignores the possibility of some indices mattering more than others, we conducted additional analyses for individual indices compared to the None condition. For each index, we tested the hypothesis that when a past event shares this index with the most recently narrated event, that past event is more salient than one which shares no indices with it. These exploratory analyses do not fully compare the relative weights of the indices and are meant primarily to suggest directions for further study.

1) Methodology: Participants read short text stories one event at a time, and were interrupted after reading a certain event (the current event) and asked to recall a certain past event. We utilized their response time to approximate salience—the shorter the response time, the more salient was the event being recalled.

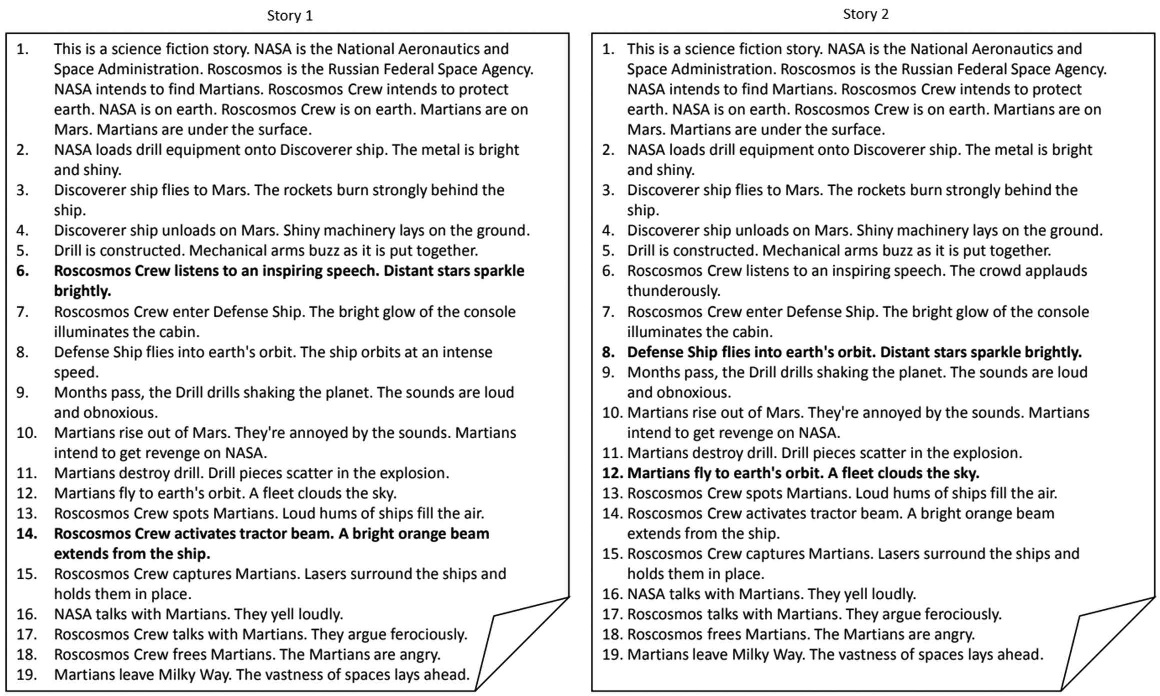
Measuring reaction time in this fashion is an established means of studying salience in memory-related tasks [46], [47]. Although we expect differences in reading time by participant as well as by story text, we did not control for this explicitly because it would require each participant to read each story multiple times. Since response times would necessarily be affected by repeated viewings, we opted to run an independent-samples design. This design accounts for individual differences by increasing the threshold for statistical significance relative to within-subjects tests. Individual differences are further assumed to be evenly distributed due to random assignment.

The null hypothesis for our primary comparison states that, among subjects who accurately remember the past event, the subject's reaction time will not differ significantly when the current and past event share any index from when they share no indices. Likewise, the null hypotheses for the individual index comparisons, which are secondary in our investigation, state that among subjects who accurately remember the past event, their reaction time will not differ significantly when the current and past event share that index from when they share no indices. Theoretically, we would presume that subjects should be faster when the current and past events share an index, acting as a memory cue to speed retrieval; however, all statistical tests are two tailed to test the possibility of slowing in the presence of shared event indices.

We designed four story domains—a zombie apocalypse, a medieval fantasy, a science fiction adventure, and a heist. In each domain, we created six versions of a simple story and selected a past and current event from each version, such that one story per domain had a past and current event that shared no indices, one had them sharing only the *protagonist* index, one the *time* index, and so on. There was one exception based on index definitions: When the past and current event share *intentionality*, they also share *protagonist* and *causality*.[1] The six versions of each story were kept as similar to each other as possible, but some variation was necessary in order to achieve the appropriate index matching between the past and current event. Examples of variations between stories are given in Fig. 1.

Since memory speed may be affected by the time that has passed since the event, we attempted to control several properties related to the length of the stories. Table I gives summary statistics for three important values. Length indicates the total

[1] Under the Indexter definition, sharing intentionality implies also sharing protagonist. 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 1

1. This is a science fiction story. NASA is the National Aeronautics and Space Administration. Roscosmos is the Russian Federal Space Agency. NASA intends to find Martians. Roscosmos Crew intends to protect earth. NASA is on earth. Roscosmos Crew is on earth. Martians are on Mars. Martians are under the surface.
2. NASA loads drill equipment onto Discoverer ship. The metal is bright and shiny.
3. Discoverer ship flies to Mars. The rockets burn strongly behind the ship.
4. Discoverer ship unloads on Mars. Shiny machinery lays on the ground.
5. Drill is constructed. Mechanical arms buzz as it is put together.
6. **Roscosmos Crew listens to an inspiring speech. Distant stars sparkle brightly.**
7. Roscosmos Crew enter Defense Ship. The bright glow of the console illuminates the cabin.
8. Defense Ship flies into earth's orbit. The ship orbits at an intense speed.
9. Months pass, the Drill drills shaking the planet. The sounds are loud and obnoxious.
10. Martians rise out of Mars. They're annoyed by the sounds. Martians intend to get revenge on NASA.
11. Martians destroy drill. Drill pieces scatter in the explosion.
12. Martians fly to earth's orbit. A fleet clouds the sky.
13. Roscosmos Crew spots Martians. Loud hums of ships fill the air.
14. **Roscosmos Crew activates tractor beam. A bright orange beam extends from the ship.**
15. Roscosmos Crew captures Martians. Lasers surround the ships and holds them in place.
16. NASA talks with Martians. They yell loudly.
17. Roscosmos Crew talks with Martians. They argue ferociously.
18. Roscosmos Crew frees Martians. The Martians are angry.
19. Martians leave Milky Way. The vastness of spaces lays ahead.

Story 2

1. This is a science fiction story. NASA is the National Aeronautics and Space Administration. Roscosmos is the Russian Federal Space Agency. NASA intends to find Martians. Roscosmos Crew intends to protect earth. NASA is on earth. Roscosmos Crew is on earth. Martians are on Mars. Martians are under the surface.
2. NASA loads drill equipment onto Discoverer ship. The metal is bright and shiny.
3. Discoverer ship flies to Mars. The rockets burn strongly behind the ship.
4. Discoverer ship unloads on Mars. Shiny machinery lays on the ground.
5. Drill is constructed. Mechanical arms buzz as it is put together.
6. Roscosmos Crew listens to an inspiring speech. The crowd applauds thunderously.
7. Roscosmos Crew enter Defense Ship. The bright glow of the console illuminates the cabin.
8. **Defense Ship flies into earth's orbit. Distant stars sparkle brightly.**
9. Months pass, the Drill drills shaking the planet. The sounds are loud and obnoxious.
10. Martians rise out of Mars. They're annoyed by the sounds. Martians intend to get revenge on NASA.
11. Martians destroy drill. Drill pieces scatter in the explosion.
12. **Martians fly to earth's orbit. A fleet clouds the sky.**
13. Roscosmos Crew spots Martians. Loud hums of ships fill the air.
14. Roscosmos Crew activates tractor beam. A bright orange beam extends from the ship.
15. Roscosmos Crew captures Martians. Lasers surround the ships and holds them in place.
16. NASA talks with Martians. They yell loudly.
17. Roscosmos talks with Martians. They argue ferociously.
18. Roscosmos frees Martians. The Martians are angry.
19. Martians leave Milky Way. The vastness of spaces lays ahead.

Fig. 1. Two versions of the Science Fiction story. In Story 1, the current and past events (14 and 6) share the protagonist index. In Story 2, the events (12 and 8) share the space index.

TABLE I
SUMMARY STATISTICS FOR VARIOUS STORY LENGTH PROPERTIES, GIVEN AS NUMBER OF EVENTS

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

number of events in the story (where the description of the initial state is considered the first event). Prompt distance is the number of events read before the subject is interrupted, and Prompt gap is the number of events between the past and current event.

One design challenge for this experiment was the need to probe the subject's memory of a past event without affecting their mental model. Simply asking if the subject remembers "Roscosmos Crew listens to an inspiring speech," for example, likely primes their response by mentioning the Roscosmos Crew characters. It was therefore necessary to attach some additional information to each event that could be used to reference the event later.

To accomplish this, we expressed each event in two sentences: First, the event itself—a text representation of a planning operator translated into simple natural language; and second, a hand-written piece of flavor text associated with the event. For example, the second event in the stories in Fig. 1 is "NASA loads drill equipment onto Discoverer ship," and the flavor text associated with that event is "The metal is bright and shiny."

The first event in each story is a complete description of the initial state. There are two special cases for which other events have additional text. When an event motivates a new character goal, this is indicated with an additional sentence at the end. For example, event 10 in Story 2 motivates the Martians to adopt the goal to get revenge on NASA. Additionally, when the time frame changes between two events, such as between events 8 and 9, this change is indicated at the beginning of the second event.

In this particular story (Fig. 1, Story 2), the past and current event (bolded) are events 8 and 12, respectively. Subjects were interrupted after reading event 12 and asked if they remember "Distant stars sparkle brightly"—the flavor text of the past event. This indirect probing of memory has been used in similar studies [48]. To ensure that the text successfully conveys which events do and do not share indices, three raters tagged the past and current events 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 $\alpha = 1$).

Indexter assumes that the audience will segment the narrative's events, time, and space into discrete units. It is important for the granularity of the segmentation scheme used by Indexter to match that of the audience's mental model as closely as possible. In an attempt to ensure this, we had participants play an initial training game containing short text events with a similar format (i.e., template sentence+flavor text). Panels at the bottom of the screen indicated the time frame (e.g., Day 1) and location (e.g., Mars) of the event currently displayed. The game was meant to prime the audience on how events, time, and space are segmented in the test stories, and to ensure that readers perceive changes in time and space when we intend them to. Previous studies have primed segmentation through visual aids of spatial arrangement [49] and through passive viewing of films prior to an experiment [50].

Stories were read online in a web browser. Participants placed two fingers from the same hand on two keys (e.g., "1" and "2") and placed the other hand on the spacebar. One event is shown at a time, and the spacebar advances to the next event. When the participant is interrupted after reading the prompt event, they are shown the flavor text from the past event and asked to press 1 for "yes" or 2 for "no" to indicate whether they remember it. The time it takes them to answer is recorded using Javascript, which has been shown to be sufficiently accurate for measuring response time across different systems and browsers [51].

Participants were asked to answer in under 2 s, but longer responses were still recorded. Different keyboard configurations were available for left and right handed participants. Prior to starting the study, participants practiced with the interface on two stories and repeated them until they were able to answer the prompts correctly in under 2 s.

This study required many participants because very little data could be collected per person. Because the interruption itself modifies the participant's mental model, only one data point can be obtained per story per user. Furthermore, we had to include some prompts whose correct answer was "no" (i.e., the prompt asks about an event which has not happened) to stop participants from believing that "yes" is always correct. After completing the training exercises, each subject read four stories—one from each of the four domains—testing a particular index (or the None case). For example, a single participant, assigned the *protagonist* index, would read the version of each story in which the designated current and past events share exactly the protagonist index. Stories were presented in random order to control for an ordering effect. Two stories had prompts for which the correct answer was "yes," and the other two, "no." Subjects were told that they would only receive compensation if they answered at least three of the four prompts correctly, though in truth participants with lower accuracy were also compensated.

Subjects were recruited via the Amazon Mechanical Turk crowdsourcing 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 were 18 years of age or older. 200 participants completed the study on Mechanical Turk, resulting in 800 responses across the 24 stories. We observed considerable variance in reaction times and a surprisingly low accuracy of only 71%. A d' analysis [52] 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.

2) Results: Our data showed no effect of index condition on retrieval accuracy ($F_5, 209 = 1.7532, p = 0.124$), but this should be investigated with more stories per individual in follow-up research. We are primarily looking at the speed of access for successful retrieval, i.e., participants' response time when recalling correct information. We 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).

In addition, due to the high variance in response time and because each participant read all four stories of the same index, we identified and removed outliers using a recursive form of the Grubbs test [53]. The recursive form has been demonstrated to be better than the classic Grubbs test for identifying multiple outliers [54]. The process was to iterate through each condition and test for the presence of an outlier in either extreme using a two-tailed test at the $p < 0.5$ level (0.025 at each tail). Significant outliers were removed and the test was repeated until no more outliers were detected. This resulted in the removal of 24 of 800 data points. A summary of the remaining 324 observations is given in Table II, broken down by index.

TABLE II
NUMBER OF OBSERVATIONS FOR EACH INDEX, ALONG WITH THE MEAN RESPONSE TIME AND STANDARD DEVIATION FOR THAT INDEX IN MILLISECONDS

Index	Count	Mean (ms)	Std. Deviation (ms)
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 III
p VALUES FOR EXPLORATORY PAIRED T-TESTS OF EACH INDEX, UNADJUSTED

	None	Protagonist	Time	Space	Causality
Protagonist	**.004**				
Time	**.021**	.612			
Space	**.022**	.553	.941		
Causality	.092	.256	.535	.577	
Intentionality	**.026**	.573	.951	.991	.580

Significant values ($p < 0.05$) are in bold.

Regression analyses revealed no significant differences in response time by prompt distance ($F_{1,320} = 0.003, p = 0.96$) or gap ($F_{1,320} = 0.428, p = 0.51$), or their interaction ($F_{1,320} = 0.087, p = 0.77$). This suggests that we succeeded in controlling information load between stories.

An independent samples t-test compared response time in trials with no matching index to trials with any matching index. Response times were found to be significantly shorter in the presence of any index (no index = 1803 ms, any index = 1559 ms; $t_{74} = 2.559, p = 0.0126$). This suggests that there is an overall effect of shared index on recall speed for an event. We reject our null hypothesis and note the directionality; in the presence of a matching index, participants remembered things faster, as expected. Accuracy did not significantly differ with the presence of any index (no index = 0.75, any index = 0.70; $t_{74} = 1.240, p = 0.217$).

Further exploratory analyses tested whether indices contributed to this speedup to differing degrees. We performed a one-way ANOVA using a dummy factor coding with None set as the 0th dummy factor, which revealed marginally significant differences in response time by index ($F_{5,318} = 2.079, p = 0.0677$). Paired t-tests compared each index to the None condition and found four of the five indices to have a significant individual effect on response time. The fifth, causality, had a marginally significant effect. The results of these tests are shown in the first column of Table III. These and their adjusted values in Table IV are intended to be taken with a grain of salt and should not be interpreted as a full comparison of the index weights.

TABLE IV
p VALUES FOR EXPLORATORY PAIRED T-TESTS OF EACH INDEX, ADJUSTED USING THE BENJAMINI AND HOCHBERG (1995) [55] CORRECTION METHOD

	None
Protagonist	**.020**
Time	**.033**
Space	**.033**
Causality	.092
Intentionality	**.033**

Significant values ($P<0.05$) are in bold.

The remaining columns in Table III show the results of the index-to-index tests, which are not part of our hypotheses. We included them only because it is interesting to see that there were no significant interactions between the indices. Table IV shows the index-to-None p-values after correction using the Benjamini & Hochberg [55] method, but we are only correcting the p-values about which we have stated hypotheses. To correct every p-value in Table III would be an overly harsh correction. All significant p-values remained significant after correction, and the marginally significant one remained marginally significant.

Participants who accurately remembered the past event responded faster when the most recently narrated event shared any index with the past event. This study was insufficiently powered to address whether some indices have a stronger effect than others (as seen in Table III), or whether sharing multiple indices has a stronger effect than sharing one index. 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, which we will address again in Section V.

B. Predicting Future Choices Based on Past Event Salience

At this point we have supported the first claim mentioned in Section I—that we can approximate the salience of past events using Indexter. To address our second claim, we demonstrate that we can use the salience model evaluated in the previous section to predict readers' choices for the final event in a story.

We speculated that readers would be more likely to choose endings that make past events more salient. Under the pairwise event salience model, this can be represented by the number of indices that an ending event shares, in total, with previous story events. This section describes our second study [11], in which we showed participants an interactive story with two possible endings and tested whether they significantly preferred the one that shared more indices with past events.

For this study, we designed an interactive story whose events could be generated by a narrative planner, such as Glaive [36] or IMPRACTical [56]. The story contained two possible endings. Before reaching the ending, readers were given four preliminary choices, each with two options. In all cases, one option shared an index with one ending, and the other shared the same index with the other ending. For the final experimental choice, we hypothesized that readers would choose whichever ending shared more indices with past events—which is to say, the ending that

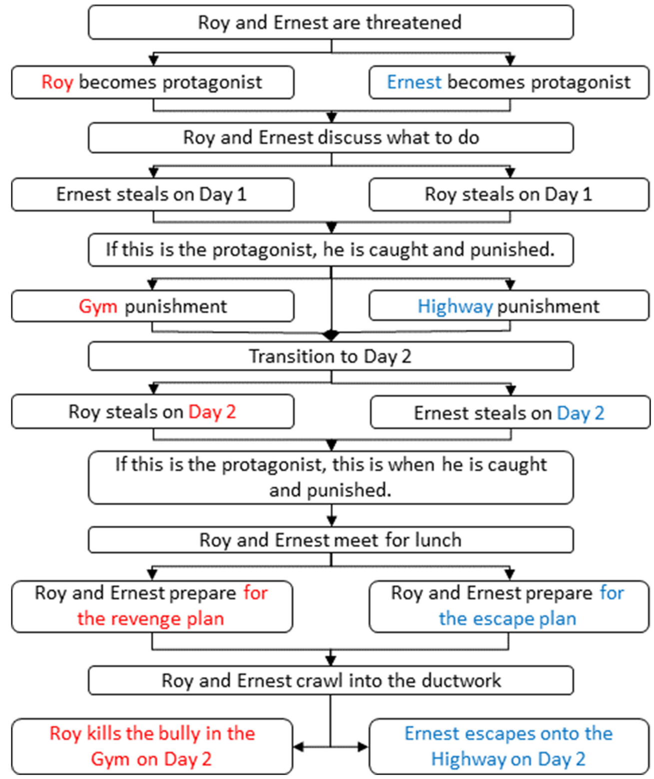


Fig. 2. Prediction study choices.

matched the majority of their previous four choices. If both endings matched exactly two of these, we made no prediction.[2]

1) Methodology: Fig. 2 shows a summary of the story, starting at the top and ending with one of the two final events at the bottom, which we will henceforth refer to as the Revenge ending and the Escape ending. Wherever there are choices, the "Revenge-salient" option is displayed on the left and the "Escape-salient" option on the right. All choices are independent of each other, with the exception of the Time choice: If Ernest is chosen to steal on Day 1, then Roy will be the one to steal on Day 2, and *vice versa*. All of the remaining studies presented here used some version of this story. We attempted to keep the stories as similar as possible across the different studies, changing only what was necessary to meet their design criteria.

The story is about two prisoners, Roy and Ernest, who are threatened to be killed by the prison bully. They each devise a plan in response. Ernest plans to break out of prison and escape onto the highway, while Roy plans to get revenge by killing the bully in the gym. Both plans involve stealing an item and then crawling into the ductwork through a loose vent. The characters decide to work together in pursuit of both plans.

In all possible versions of the story, Roy and Ernest end up inside the ductwork, ready to complete either of the two plans,

[2] We chose not to include a choice based on the *causality* index because of the following complication: If a choice toggles between two events, each of which is causally related to only one ending, then only one ending will be enabled after this choice is made. However, it is essential in our experiment that both endings are possible in all versions of the story.

when a guard discovers their whereabouts at the last minute. However, the guard believes there is only *one* prisoner in the duct rather than two. Roy and Ernest realize that if they continue together, they will both be caught and also, the goal will not be accomplished; but if one of them turns himself in, the other will have time to complete his original plan. The reader is then prompted to choose between the "Escape" ending (where Ernest escapes onto the highway) and the "Revenge" ending (where Roy kills the bully in the gym). The following is a description of how we manipulate each Indexter index before arriving at this experimental choice.

> Choice 1 (protagonist) prompts the reader to choose whether it is Roy or Ernest who takes a piece of contraband. This will be the character who later gets caught and sent to a punishment after performing his theft, constituting an additional scene in which this character is the protagonist. This becomes an additional event that can be made salient by this character's ending, but not the other ending. We say that Choice 1 gives one vote to the chosen character's ending: Choosing Roy is a vote for his Revenge ending, and choosing Ernest is a vote for his Escape ending.
>
> Choice 2 (time) determines in what order the characters perform their thefts. The vote goes to the character who performs his theft on Day 2, because the ending also happens on Day 2 and thus shares the time index with the second theft.
>
> Choice 3 (space) prompts the reader to choose between two punishments—one located on the highway like the Escape ending, the other in the gym like the Revenge ending. The vote goes to the ending sharing the chosen location.
>
> Choice 4 (intentionality) gives two options for a final action that the characters perform together—they either lock the bully in the gym (for the same goal as the Revenge ending), or they put on their civilian disguises (for the same goal as the Escape ending). These actions are a necessary part of their respective plans, but can be performed *either* before entering the ductwork or after exiting it. In other words, not performing one of these actions at this point does not render its goal unachievable.

To summarize, when the participant reached the final choice between the two endings, we hypothesized that they would choose the ending for which more of the following were true (we made no predictions if the endings were tied):

1) its character is the same as the character who had one extra scene (protagonist);
2) its location is the same as the location of the punishment scene (space);
3) its character is the same as the character who stole his item on Day 2 (time);
4) its goal is the same as the goal of the preparatory action (intentionality).

2) Results: We manually translated the stories into natural language. We built the story using Twine, an opensource tool for writing branching stories in HTML. Events were displayed one at a time in short passages, along with a graphic depicting the character, location, and time (Day 1 or Day 2) of the current scene. After a certain amount of time, a link to the next passage appears (or multiple, if there is a choice). An example of a passage is shown in Fig. 3.

We recruited 350 participants through Amazon Mechanical Turk, and paid them each \$0.25 for completing the story. The

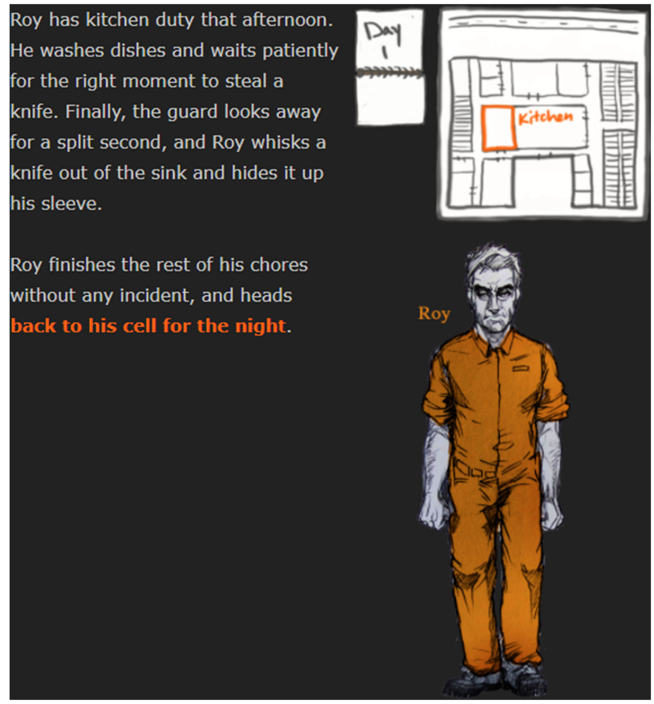


Fig. 3. Example twine passage.

TABLE V
PREDICTING CHOICES—RESULTS

	Chose Escape	Chose Revenge
Expected Escape	32	14
Expected Revenge	11	21

$p < 0.0022$.

average time spent on the study was 7.5 min. In order to filter out the participants who were not fully reading the story, we asked a series of comprehension questions after the story was completed. The questions were designed to verify that the pertinent information was accurately communicated to the reader. We discarded the data from participants who were unable to answer all the questions correctly. Those who answered them all correctly were awarded a \$0.75 bonus, the availability of which was made known to all participants at the start.

Of the 350 results, we discarded 225 and were left with 125 responses from participants who demonstrated full comprehension of the story. Of these 125 results, there were 78 for which a majority of the reader' s choices were in favor of one ending or the other. (For the others, we made no prediction.) We conducted our evaluation using the remaining 78 results.

To evaluate our hypothesis we used Fisher' s exact test, which is similar to the $\chi 2$ test, but performs better for distributions with small expected values [57]. Fisher' s exact test is nonparametric, meaning it does not assume any underlying distribution of the population. This is important because participants chose more Escape options overall than Revenge ones (perhaps due to the morality differences between the two paths). Fisher' s exact test is not skewed by this imbalance.

TABLE VI
INFLUENCING CHOICES—VERSION A RESULTS

	Chose Escape	Chose Revenge
Expected Escape	17	15
Expected Revenge	22	14

$p = 0.8286$.

TABLE VII
INFLUENCING CHOICES—VERSION B RESULTS

	Chose Escape	Chose Revenge
Expectd Escape	27	21
Expected Revenge	26	14

$p = 0.854$.

TABLE VIII
INFLUENCING CHOICES—VERSION C RESULTS

	Chose Escape	Chose Revenge
Expected Escape	22	14
Expected Revenge	25	13

$p = 0.7451$.

Table V shows the contingency table giving the frequency distribution of results according to their expected outcomes. The null hypothesis that the ending choices were independent of the Indexter indices of previous events, was rejected with $p < 0.0022$.

The odds ratio for this contingency table is ≈ 4.27, meaning there are about 4 to 1 odds that participants chose the ending we expected them to choose. We concluded that readers were indeed more likely to choose future events which would make past events more salient.

C. Influencing Future Choices Based on Past Event Salience

To support our third and final claim, we wanted to show that we could not only *predict* readers' choices for final events, but actually *influence* them to choose whichever ending we selected. This meant that we could no longer allow the readers' preliminary choices to determine which ending we predicted for them—our prediction must be made from the beginning. Using variations of the prisoners story, we conducted a series of increasingly informative experiments to try to achieve this.

1) No Choices Version: The simplest modification of the prediction study that would allow us to attempt to influence readers' ending choices would be to predetermine the four preliminary events rather than allowing readers to choose them. For example, we should be able to influence readers to choose the Escape ending simply by selecting all of the choices that make the Escape ending more salient.

We created two versions of the story in this manner; one in which we selected the Escape-salient option for all four preliminary choices, and one in which we selected all four Revenge-salient options. We removed the prompts for those choices and simply narrated the resulting linear story. In these versions, the readers made no choices other than the experimental choice at the end.

a) Results: We divided participants into two groups; one for each story version. Using the same experimental setup as the prediction study, we had 32 participants read the Escape-salient version and correctly answer validation questions, and 36 for the Revenge-salient version. The results are presented in Table VI. In this version, Fisher' s exact test failed to reject the null hypothesis, with $p = 0.8286$.

2) Unrelated Choices Version: Since this failed version contained no choices other than the final choice for the ending, it is possible that the results were skewed by the lack of interactivity of the story compared to the version used in the prediction study. To rule out this possibility, we conducted another version of the influence study which was exactly the same as the previous one, but with the addition of four choices that were unrelated to the salience of the endings. This included choices, such as "Where does Roy hide the knife: In his pant leg, or in his sleeve?" As with the previous version, we divided subjects into an all-Revenge group and an all-Escape group and attempted to influence readers in each group to choose the associated ending.

a) Results: We had 48 results from the Escape version, and 40 for the Revenge version. The results are shown in Table VII. Fisher' s exact test failed once again to reject our null hypothesis, with $p = 0.854$. We conclude that the failure of the previous version was *not* simply due to the lack of interactivity in the story.

Instead, we believe it was due to the lack of choices specifically related to the endings. In the prediction study, the past events that were being made salient by the endings were direct results of the reader's choices, but this is no longer the case. We conclude that in order to successfully influence readers by manipulating event indices, we must consider specifically the indices of those events that the reader chose to happen. The next approach builds on this insight.

3) Low Agency Choices Version: In the next version, we engineered the story so that the readers once again chose the four key events, except this time they were given two options which both made the same ending salient (rather than one for each). For example, when participants in the Escape group were prompted to choose between two punishments, their options were picking up trash along the highway, or mowing the grass along the highway. In either case, the location of the resulting event is the Highway, and thus it will share the space index with the Escape ending regardless of which option they choose. Similarly, participants in the Revenge group were given a choice between two punishment duties that both take place in the gym.

a) Results: We again created two versions of the story where all four choices were handled in this manner; either all sharing indices with the Escape ending, or all sharing indices with the Revenge ending. We had 36 valid responses from the Escape group and 38 from the Revenge group. As shown in Table VIII, we were once again unsuccessful in influencing readers' final choices. Fisher' s exact test fails to reject our null hypothesis, with $p = 0.7451$.

From this, we can make an important observation. In the prediction study, each choice featured two options that *differed* along at least one index. In this study, however, the two options always have the same values for all indices.

A recent study [41] using Indexter showed that when players were given options that differed along at least one EISM index, they reported feeling more agency than when their options differed along no indices. This suggests that the success of the prediction study may have been contingent upon the choices feeling *meaningful*. A better conclusion might be that when readers are given meaningful choices, they have a tendency to choose endings that remind them of those choices.

4) High Agency Choices Version: In our final version of this study [12], we aimed to influence readers once again by reminding them of their past choices, but this time all of those choices were "high agency" choices as described in the previous section. That is, the options presented were always different from each other in at least one index, as was the case in the original prediction study.

To account for this new constraint, we modified the story so that each choice toggled between one option that shared an index with the targeted ending, and one option that did not share that index with either of the two endings. Therefore the targeted ending could receive anywhere from 0 to 4 votes, while the other ending could receive none. If the targeted ending has at least 1 vote when the final choice is reached, this means it has more indices in common with the reader's previous (meaningful) choices than the other ending does, and therefore we expect them to choose the one we targeted.

The following is a description of how we modified each of the four choices.

> Choice 1 (Protagonist): In this version we introduce a friendly guard Mitchell, who warns Roy and Ernest of their death threat and gives them his key card to help them. For the protagonist choice, instead of choosing between Roy and Ernest, participants in the Revenge group choose between Roy and Mitchell, while those in the Escape group choose between Ernest and Mitchell. Since Mitchell is not present in either of the two endings, neither ending will share the protagonist index with this event if Mitchell is selected.
>
> Choice 2 (Time): Rather than allowing this choice to determine when both characters perform their thefts, we fix the "other" character's theft to always happen on Day 1, the nonsalient day. Now this choice simply determines on which day the targeted ending's character performs his theft—either on Day 2, which gives that ending a vote, or also on Day 1 (i.e., both thefts happen on the first day), which gives neither ending a vote.
>
> Choice 3 (Space): We introduced an additional punishment option: Cleaning the bathroom. Readers choose between the punishment in the bathroom or the punishment in the same location as the targeted ending for their group.
>
> Choice 4 (Intentionality): Readers can choose between two actions: Either the action used in the prediction study that shares its goal with the targeted ending, or a new action that does not share intentionality with either ending—returning the key card to Mitchell. This is an additional goal that both characters have, which was introduced in an earlier scene.

We expected that participants in the Escape group would choose the Escape ending, and participants in the Revenge group would choose the Revenge ending, except in cases where the participant made none of the choices that give votes to the targeted ending. In those cases the two endings shared an equal number of indices with past choices, so we have no prediction to make.

TABLE IX
INFLUENCING CHOICES RESULTS (≥ 1 VOTE)

	Chose Escape	Chose Revenge
Expected Escape	46	16
Expected Revenge	32	30

$p < 0.0076$, odds ratio: 2.67.

TABLE X
INFLUENCING CHOICES RESULTS (≥ 2 VOTES)

	Chose Escape	Chose Revenge
Expected Escape	30	10
Expected Revenge	15	14

$p < 0.04$, odds ratio: 2.76.

TABLE XI
INFLUENCING CHOICES RESULTS (≥ 3 VOTES)

	Chose Escape	Chose Revenge
Expected Escape	12	3
Expected Revenge	5	5

$p = 0.128$, odds ratio: 3.76.

a) Results: Using the same method as previous studies, we received 124 valid responses in which the participant made at least one choice that gave a vote to the targeted ending. Fisher' s exact test (Table IX) rejected the null hypothesis with $p < 0.0076$.

We further analyzed the subset of these 124 responses in which the reader made at least *two* choices in favor of the targeted ending, and again for those who made at least three. There was only one response in which the reader made all four of the choices in favor of the targeted ending, so we were unable to evaluate the fourth table. We expected to see a stronger preference for the targeted ending with each subset. Table X shows the frequency distributions for the 69 responses in which the reader made at least two choices in favor of the targeted ending.

In this subset, the null hypothesis was rejected again ($p < 0.04$) and the odds ratio increased to 2.76. Table XI shows the 25 responses in which the reader made at least three choices in favor of the targeted ending.

The third set was insufficiently powered to reject the null hypothesis ($p = 0.128$), but we speculate that it would become significant with a larger sample size.

To avoid a multiple comparisons problem and in attempt to investigate whether the apparent increase in odds ratios over these tables is statistically significant, we conducted a single test on the full set of data (from Table IX) using logistic regression and two terms: Our prediction, and the number of votes (1, 2, 3, or 4). The test confirmed the significance of our prediction's effect ($p < 0.0131$), but showed no significant effect from the number of votes. Therefore we cannot confirm any effect from the *number* of past choices being made salient by the ending; only from the presence or absence of such choices.

V. DISCUSSION

In this paper, we first demonstrated that the Indexter model can effectively approximate the salience of past events in the audience's memory. Using Indexter's model of mapping the five

EISM indices to narrative planning events, we can define when two events share each index, and we can say that a past event becomes more salient when it shares at least one index with the current (or most recently narrated) event.

Next, we showed that we can use this model of salience to predict players' choices for ending events. When presented with two ending options, readers were significantly more likely to choose the ending that shared more indices with previous events they had chosen in the past. Finally, we showed that we can use the same method to *influence* players' choices for ending events, provided that we ensure some degree of agency in the important past choices.

It is important to clarify how salience has been isolated in these experiments. It may seem a plausible alternative explanation that readers are predisposed by their personalities to choose certain types of conflict resolution (e.g., peaceful versus violent), and that these dispositions are contributing to their choices throughout the story. However, we designed the studies in Section IV-C to avoid this potentially confounding factor. Although personal dispositions may have led readers to make more or fewer choices consistent with their targeted ending, this number did not affect the prediction we made for their final choice. Our predictions were based solely on their grouping, determined at the onset of the experiment.

Future work can improve the pairwise model of event salience in several ways. As mentioned in an earlier section, we are assuming that all indices are equally weighted simply because we have no empirical data suggesting what their individual weights should be. We are also ignoring the potential effects of two events sharing *multiple* indices—we only consider whether or not they share at least one index. In addition, we ignore the salience generated by the *second*-most-recently narrated event, and the event before that, and so on. While this may be sufficient for short stories, longer stories would benefit from a more robust model that takes into account how salience decays over time.

The specific conclusions we made in this work regarding predicting and influencing choices may only pertain to endings. More work needs to be done to test whether this is a special case, or whether the same method would work to predict choices throughout different points in the story. It is possible, for example, that we would find the opposite to be true during the beginning stages of a story—that readers may prefer events that take them to new places, introduce new characters, and so on, rather than reminding them of things that have already happened.

While the particular method used here to achieve influence may be narrow, the broader claim we are supporting is that an effective model of salience and memory can be useful toward modeling the player's preferences and desires for the future. Given that Indexter provides a simple model for measuring salience that is easy to implement on top of plan structures, we believe that this is a promising tool for plan-based interactive narrative systems.

Acknowledgment

The authors would like to acknowledge Christopher Kives for his contributions to Study 1; the original authors of Indexter: Rogelio E. Cardona-Rivera, Bradley A. Cassell, and R. Michael Young; and artist Caroline Catlett Gates for the character drawings used in Studies 2 and 3.

References

[1] M. O. Riedl, A. Stern, D. Dini, and J. Alderman, "Dynamic experience management in virtual worlds for entertainment, education, and training," *Int. Trans. Syst. Sci. Appl. Special Issue Agent Based Syst. Human Learn.*, vol. 4, no. 2, pp. 23–42, 2008.

[2] B. Magerko, "Player modeling in the interactive drama architecture," Ph.D. dissertation, Dept. Comput. Sci. Eng., Univ. Michigan, Ann Arbor, MI, USA, 2006.

[3] M. Sharma, M. Mehta, and S. Ontanon, "Player modeling evaluation for interactive fiction," in *Proc. 3rd Artif. Intell. Interactive Digit. Entertainment Conf.*, 2007, pp. 19–24.

[4] H. Barber and D. Kudenko, "Dynamic generation of dilemma-based interactive narratives," in *Proc. 3rd AAAI Conf. Artif. Intell. Interactive Digit. Entertainment*, 2007, pp. 2–7.

[5] D. Thue, V. Bulitko, M. Spetch, and E. Wasylishen, "Interactive storytelling: A player modelling approach." in *Proc. 3rd AAAI Int. Conf. Artif. Intell. Interactive Digit. Entertainment*, 2007, pp. 43–48.

[6] R. Farrell, S. Robertson, and S. G. Ware, "Asking hypothetical questions about stories using quest," in *Proc. Interactive Storytelling: 9th Int. Conf. Interactive Digit. Storytelling*, Los Angeles, CA, USA, 2016, pp. 136–146.

[7] R. E. Cardona-Rivera, B. A. Cassell, S. G. Ware, and R. M. Young, "Indexter: A computational model of the event-indexing situation model for characterizing narratives," in *Proc. 3rd Workshop Comput. Models Narrative*, 2012, pp. 34–43.

[8] R. A. Zwaan and G. A. Radvansky, "Situation models in language comprehension and memory," *Psychological Bull.*, vol. 123, no. 2, pp. 162–185, 1998.

[9] R. E. Cardona-Rivera and R. M. Young, "A knowledge representation that models memory in narrative comprehension," in *Proc. 28th AAAI Conf. Artif. Intell—Student Abstr. Track*, 2014, pp. 3098–3099. [Online]. Available: https://liquidnarrative.csc.ncsu.edu/wp-content/uploads/sites/15/2015/11/cardona-rivera2014knowledge.pdf

[10] C. Kives, S. G. Ware, and L. J. Baker, "Evaluating the pairwise event salience hypothesis in Indexter," in *Proc. 11th AAAI Int. Conf. Artif. Intell. Interactive Digit. Entertainment*, 2015, pp. 30–36.

[11] R. Farrell and S. G. Ware, "Predicting user choices in interactive narratives using indexter's pairwise event salience hypothesis," in *Proc. 9th Int. Conf. Interactive Digit. Storytelling*, 2016, pp. 147–155.

[12] R. Farrell and S. G. Ware, "Influencing user choices in interactive narratives using indexter's pairwise event salience hypothesis," in *Proc. 13th AAAI Int. Conf. Artif. Intell. Interactive Digit. Entertainment*, 2017, pp. 37–42.

[13] D. C. Dennett, *The Intentional Stance*. Cambridge, MA, USA: MIT Press, 1989.

[14] J. Bates *et al.*, "The role of emotion in believable agents," *Commun. ACM*, vol. 37, no. 7, pp. 122–125, 1994.

[15] A. B. Loyall, "Believable agents: Building interactive personalities," Ph.D. dissertation, Dept. Comput. Sci., Carnegie-Mellon Univ., Pittsburgh, PA, USA, 1997.

[16] M. Riedl, C. J. Saretto, and R. M. Young, "Managing interaction between users and agents in a multi-agent storytelling environment," in *Proc. 2nd Int. Joint Conf. Auton. Agents Multiagent Syst.*, 2003, pp. 741–748.

[17] M. Mateas and A. Stern, "Structuring content in the façade interactive drama architecture," in *Proc. 1st AAAI Int. Conf. Artif. Intell. Interactive Digit. Entertainment*, 2005, pp. 93–98.

[18] B. Laurel, "Toward the design of a computer-based interactive fantasy system," Ph.D. dissertation, Dept. Theatre, Ohio State University, Columbus, OH, USA, 1986.

[19] P. W. Weyhrauch, "Guiding interactive drama," Ph.D. dissertation, School Comput. Sci., Carnegie Mellon Univ., Pittsburgh, PA, USA, 1997.

[20] M. Mateas and A. Stern, "Integrating plot, character and natural language processing in the interactive drama façade," in *Proc. 1st Int. Conf. Technol. Interactive Digit. Storytelling Entertainment*, vol. 2, 2003.

[21] R. M. Young, M. O. Riedl, M. Branly, A. Jhala, R. Martin, and C. Saretto, "An architecture for integrating plan-based behavior generation with interactive game environments," *J. Game Develop.*, vol. 1, no. 1, pp. 51–70, 2004.

[22] M. J. Nelson and M. Mateas, "Search-based drama management in the interactive fiction anchorhead," in *Proc. 1st AAAI Conf. Artif. Intell. Interactive Digit. Entertainment*, 2005, pp. 99–104.

[23] M. J. Nelson, M. Mateas, D. L. Roberts, and C. L. Isbell, "Declarative optimization-based drama management in interactive fiction," *IEEE Comput. Graph. Appl.*, vol. 26, no. 3, pp. 32–41, May/Jun. 2006.

[24] H. Yu and M. O. Riedl, "Personalized interactive narratives via sequential recommendation of plot points," *IEEE Trans. Comput. Intell. AI Games*, vol. 6, no. 2, pp. 174–187, Jun. 2014.

[25] D. L. Roberts and C. L. Isbell, "Lessons on using computationally generated influence for shaping narrative experiences," *IEEE Trans. Comput. Intell. AI Games*, vol. 6, no. 2, pp. 188–202, Jun. 2014.

[26] M. S. El-Nasr, A. V. Vasilakos, C. Rao, and J. A. Zupko, "Dynamic intelligent lighting for directing visual attention in interactive 3-d scenes," *IEEE Trans. Comput. Intell. AI Games*, vol. 1, no. 2, pp. 145–153, Jun. 2009. [Online]. Available: http://dx.doi.org/10.1109/TCIAIG.2009.2024532

[27] J. P. Rowe and J. C. Lester, "Modeling user knowledge with dynamic bayesian networks in interactive narrative environments," in *Proc. AAAI Int. Conf. Artif. Intell. Interactive Digit. Entertainment*, 2010, pp. 57–62.

[28] A. C. Graesser, S. E. Gordon, and L. E. Brainerd, "QUEST: A model of question answering," *Comput. Math. Appl.*, vol. 23, no. 6, pp. 733–745, 1992.

[29] D. B. Christian and R. M. Young, "Comparing cognitive and computational models of narrative structure," in *Proc. 19th Nat. Conf. Amer. Assoc. Artif. Intell.*, 2004, pp. 385–390.

[30] R. E. Cardona-Rivera, T. Price, D. Winer, and R. M. Young, "Question answering in the context of stories generated by computers," *Advances Cogn. Syst.*, vol. 4, pp. 227–245, 2016.

[31] R. M. Young and R. E. Cardona-Rivera, "Approaching a player model of game story comprehension through affordance in interactive narrative," in *Proc. 4th Workshop Intell. Narrative Technol.*, 2011, pp. 123–130.

[32] K. Johnstone, *Impro: Improvisation and the Theatre*. Evanston, IL, USA: Routledge, 2012.

[33] Z. Tomaszewski, "On the use of reincorporation in interactive drama," in *Proc. 18th AIIDE Conf. Intell. Narrative Technol. IV*, 2011, pp. 84–91.

[34] R. M. Young, S. G. Ware, B. A. Cassell, and J. Robertson, "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 Formal Comput. Models Narrative*, vol. 37, no. 1–2, pp. 41–64, 2013.

[35] M. O. Riedl and R. M. Young, "Narrative planning: Balancing plot and character," *J. Artif. Intell. Res.*, vol. 39, no. 1, pp. 217–268, 2010.

[36] S. G. Ware and R. M. Young, "Glaive: A state-space narrative planner supporting intentionality and conflict," in *Proc. 10th AAAI Int. Conf. Artif. Intell. Interactive Digit. Entertainment*, 2014, pp. 80–86.

[37] J. Teutenberg and J. Porteous, "Incorporating global and local knowledge in intentional narrative planning," in *Proc. Int. Conf. Auton. Agents Multiagent Syst.*, 2015, pp. 1539–1546.

[38] Y.-G. Cheong and R. M. Young, "Narrative generation for suspense: Modeling and evaluation," in *Proc. 1st Int. Conf. Interactive Digit. Storytelling*, 2008, pp. 144–155.

[39] B.-C. Bae and R. M. Young, "A computational model of narrative generation for surprise arousal," *IEEE Trans. Comput. Intell. Artif. Intell. Games*, vol. 6, no. 2, pp. 131–143, Jun. 2014.

[40] A. Jhala and R. M. Young, "Cinematic visual discourse: Representation, generation, and evaluation," *IEEE Trans. Comput. Intell. Artif. Intell. Games*, vol. 2, no. 2, pp. 69–81, Jun. 2010.

[41] R. E. Cardona-Rivera, J. Robertson, S. G. Ware, B. Harrison, D. L. Roberts, and R. M. Young, "Foreseeing meaningful choices," in *Proc. 10th AAAI Conf. Artif. Intell. Interactive Digit. Entertainment*, 2014, pp. 9–15.

[42] R. E. Fikes and N. J. Nilsson, "STRIPS: A new approach to the application of theorem proving to problem solving," *Artif. Intell.*, vol. 2, no. 3, pp. 189–208, 1972.

[43] J. P. Magliano, J. Miller, and R. A. Zwaan, "Indexing space and time in film understanding," *Appl. Cogn. Psychol.*, vol. 15, no. 5, pp. 533–545, 2001.

[44] J. M. Zacks, N. K. Speer, and J. R. Reynolds, "Segmentation in reading and film comprehension," *J. Exp. Psychol.: General*, vol. 138, no. 2, pp. 307–327, 2009.

[45] D. McAllester and D. Rosenblitt, "Systematic nonlinear planning," *Proc. 9th Nat. Conf. Artif. Intelligence*, vol. 2, pp. 634–639, 1991.

[46] G. Gillund and R. M. Shiffrin, "A retrieval model for both recognition and recall," *Psychological Rev.*, vol. 91, no. 1, pp. 1–67, 1984.

[47] A. P. Yonelinas and L. L. Jacoby, "Dissociations of processes in recognition memory: Effects of interference and of response speed," *Can. J. Exp. Psychol.*, vol. 48, no. 4, pp. 516–535, 1994.

[48] W. F. Brewer and D. A. Dupree, "Use of plan schemata in the recall and recognition of goal-directed actions," *J. Exp. Psychol.: Learn., Memory, Cognition*, vol. 9, no. 1, pp. 117–129, 1983.

[49] D. G. Morrow, G. H. Bower, and S. L. Greenspan, "Updating situation models during narrative comprehension," *J. Memory Lang.*, vol. 28, no. 3, pp. 292–312, 1989.

[50] J. M. Zacks, B. Tversky, and G. Iyer, "Perceiving, remembering, and communicating structure in events," *J. Exp. Psychol.: General*, vol. 130, no. 1, pp. 29–58, 2001.

[51] S. Reimers and N. Stewart, "Presentation and response timing accuracy in adobe flash and html5/javascript web experiments," *Behav. Res. Methods*, vol. 47, no. 2, pp. 309–327, 2014.

[52] N. A. Macmillan and C. D. Creelman, *Detection Theory: A User's Guide*. New York, NY, USA: Taylor & Francis, 2004.

[53] F. E. Grubbs, "Procedures for detecting outlying observations in samples," *Technometrics*, vol. 11, no. 1, pp. 1–21, 1969.

[54] R. B Jain, "A recursive version of Grubbs' test for detecting multiple outliers in environmental and chemical data," *Clin. Biochemistry*, vol. 43, pp. 1030–1033, 2010.

[55] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *J. Roy. Statist. Soc., Series B (Methodological)*, vol. 57, pp. 289–300, 1995.

[56] J. Teutenberg and J. Porteous, "Efficient intent-based narrative generation using multiple planning agents," in *Proc. Int. Conf. Auton. Agents Multiagent Syst.*, 2013, pp. 603–610.

[57] J. Fleiss, B. Levin, and M. Paik, *Statistical Methods for Rates and Proportions* (Wiley Series in Probability and Statistics). Hoboken, NJ, USA: Wiley, 2013.

Rachelyn Farrell received the B.S. degree in computer science from the University of Mississippi, Oxford, MS, USA, in 2012, and the M.S. degree in computer science from the University of New Orleans, New Orleans, LA, USA, in 2017. She is currently working toward the Ph.D. degree at the University of New Orleans, where she is a student and a Research Assistant.

Her research interests include computational narrative, automated planning, interactive narrative systems, and virtual environments.

Stephen G. Ware (M'08) received the B.S. degree in both computer science and philosophy from Loyola University, New Orleans, LA, USA, in 2008, the M.S. degree in computer science from North Carolina State University, Raleigh, NC, USA, in 2011, and the Ph.D. degree in computer science from North Carolina State University in 2014.

Currently, he is an Assistant Professor of Computer Science and Director of the Narrative Intelligence Lab, University of New Orleans. His research interests include computer narrative, especially automated planning, knowledge representation, narratology, cognitive science, and artificial intelligence in games.

Dr. Ware is a member of the Association for the Advancement of Artificial Intelligence, the International Game Developers Association, and the Association for Computing Machinery.

Lewis J. Baker received the B.S. degree in psychological science and religious studies from Loyola University, New Orleans, LA, USA, in 2010, and the Ph.D. degree in cognitive psychology from Vanderbilt University, Nashville, TN, USA, in 2016, before completing a Postdoctoral Fellowship at the Rutgers Institute for Data Science, Learning & Applications, Newark, NJ, USA.

Currently he is the Head of data science at Pymetrics, Inc., where they develop machine learning algorithms to predict job performance using neuroscience-based assessments. He is a key contributor to audit-AI, an opensource toolkit for assessing algorithmic bias. His research interests include everyday failures of attention and the ability to reason about other minds.