communications psychology
A Nature Portfolio journal

Article

https://doi.org/10.1038/s44271-025-00243-4

Free recall is shaped by inference and scaffolded by event structure

Check for updates

Ata B. Karagoz [1] ✉, Wouter Kool[1,2] & Zachariah M. Reagh[1,2]

Though everyday life is continuous, people understand and remember experiences as discrete events separated by boundaries. Event boundaries influence the temporal structure of memory, and have been proposed to enhance encoding of boundary-adjacent information. However, the extent to which event boundaries influence memory for specific items, and their effect on memory in interactive environments are not well understood. Here, we designed a task to test how boundaries between hidden rules and uncertainty about those rules affect free recall of item-level information. Participants (n = 66) responded to a sequence of individual word stimuli, with words grouped by hidden rules forming events, and abrupt shifts between rules causing event boundaries. Afterwards, participants freely recalled words from the task. Recall was clustered based on event structure, such that words from the same discrete event tended to be recalled together. Contrary to predictions of theories of event cognition, recall was worse for words encoded immediately after event boundaries. Finally, we used a reinforcement-learning model to characterize recall performance, allowing us to infer a positive relationship between decision certainty and recall success. These findings indicate that the structure of events and inferences made over that structure play important roles in shaping episodic memories.

Imagine playing a few rounds of poker with some friends. Your prior experiences playing poker with these friends enable you to infer that one friend is bluffing, and you choose to raise to get them to fold. This cycle of inferring and interactively testing your inferences happens across many everyday experiences, such as cooking, car maintenance, and programming. Even though you experienced the game as a continuous stream of sensory input, you might remember later the discrete event where you capitalized on a "tell" from your friend. The parsing of continuous experience into these sorts of events is studied in the field of event cognition as *event segmentation*[1–4]. Event segmentation is consistent across people, even with minimal instructions, and enables us to effectively learn from and remember the world around us (see Zacks 2020[5] for a review). However, despite much progress in recent years, the consequences of event structure on memory representations, especially in interactive and inferential settings, remain unclear.

Several theoretical frameworks have been proposed to explain the way people segment and represent continuous events. Among these are Event Segmentation Theory[1] (EST), the Event Indexing Model[6], and the Event Horizon Model[4]. Here, we focus on EST given its prominent role in guiding recent studies of episodic memory. EST posits that observers construct active event models that predict incoming observations. Large deviations between prediction and observed outcomes trigger *prediction errors*, which lead to the perception of an event boundary. A present limitation of studies of event cognition is that these observations largely stem from passive viewing experiments[3,7,8] (though see Radvansky and Copeland[9]). However, in many real-world activities, people are not merely passive observers. Instead, they interact with their environment and use the information from these interactions to learn about the world. For instance, merely observing people play a hand of poker might be represented and remembered differently than if you were playing the hand yourself. Such differences could arise due to the rapidly changing nature of goals as one plays the game leading to different attentional settings[10,11]. This reveals an important gap in our understanding of event segmentation and its effects on memory: to what extent do theories such as EST and its body of supporting evidence extend to dynamic, active environments where individuals are making choices to resolve uncertainty about the events in question?

EST makes predictions about changes to both associative memory and item memory through separate processes. According to EST, differences in associative memory are thought to occur because of binding processes within events and separation processes due to boundaries between events. The coming together of items within events and separation across them has been borne out by numerous studies[2,11–13]. The theory also predicts that memory for information around boundary points will be enhanced. This enhancement is thought to occur due to a sharpening of representations at

[1]Psychological and Brain Sciences, Washington University in St. Louis, St. Louis, MO, USA. [2]These authors contributed equally: Wouter Kool, Zachariah M. Reagh.
✉e-mail: a.b.karagoz@wustl.edu

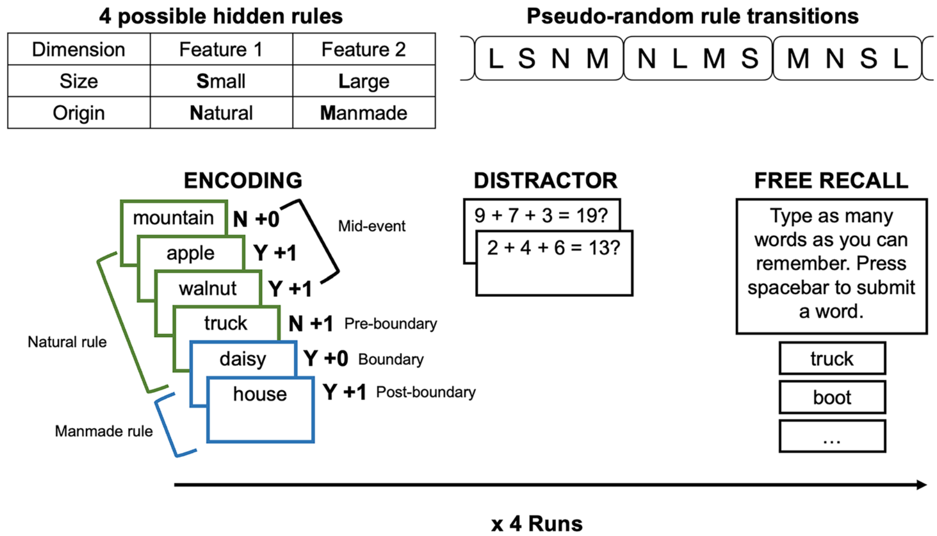

Fig. 1 | Experimental design. In the WRIT, participants indicate whether a word agrees with one of four hidden rules. The active rule repeats across events that last 6–8 trials. Then, it transitions pseudo-randomly such that each rule is experienced an equivalent number of times in each run. Here, the word "daisy" marks the change from the natural rule (green) to the manmade rule (blue) (color added as aid, not shown to participants). The capital "N" or "Y" next to each box in the ENCODING section denotes the participant's response, while the "+0" and "+1" denote received reward. After performing the word rule inference task, participants perform a distractor task for 10 seconds and are then asked to perform typed free recall. They perform this sequence 4 times with run-unique words each time.

the end of an event and the opening of perceptual gates and heightening of attention at the beginning of a new event[3] (also see Richmond and Zacks[14]; Clewett, DuBrow, and Davachi[15] for review). In short, EST argues that prediction errors during event boundaries establish the need for a new event model, and that this process leads to enhanced within-event associative memory and enhanced memory for items occurring near boundaries.

A parallel framework used to interpret event cognition is latent cause inference[16]. Here, people are thought to generate and test hypotheses about the state of the world. As in EST, prediction errors can also play a key role signaling the need to reevaluate hypotheses. This suggests an intriguing analogy between the hypotheses formed during latent cause inference and event models in EST. For example, hypotheses about the environment lead to sets of predictions that can be actively tested, which is akin to testing candidate event models[16]. Though such mechanisms have been proposed[16–21], specific experimental evidence for this claim is currently limited. The primary benefit of the latent cause inference approach is that it has explicitly tested how individuals interact with their environment and resolve uncertainty[22–24] (see Radulescu, Shin, and Niv[25] for review). Inspired by this work, we sought to bridge these parallel lines of research by asking how actively inferring latent states influences event segmentation, and how this, in turn, affects long-term memory.

To test how event segmentation and inference in an interactive environment affect the organization of episodic memory, we designed a Word Rule Inference Task (WRIT). Participants inferred an active "task rule" by interacting with the environment and receiving feedback[23,26,27]. Specifically, they indicated whether a word agreed with a hidden rule, receiving rewards when they responded correctly. This hidden rule shifted at unpredictable intervals, causing event boundaries. After this task, participants freely recalled all the words they could remember. Importantly, this final phase allowed us to examine how event boundaries caused by rule shifts and the ensuing inferential process affected the structure of recall[28].

Methods

Word rule inference task

We designed a word rule inference task (WRIT) inspired by a variant of a Wisconsin Card Sorting Task[23], which we optimized to test episodic memory. In the WRIT, participants judged whether words agreed with a hidden active rule. On each trial, the participant was shown a word and indicated if they believed the word agreed with the active rule via simple yes/no responses. They were given up to 3 seconds to respond, and the word remained on screen after their response to ensure that all words were observed for the same duration. Then, the word disappeared, and participants received binary feedback in the form of points displayed for 1 second. This feedback, in combination with the characteristics of the last word, allowed participants to deduce the active rule. Critically, the active rule changed every 6–8 trials, but this was not explicitly signaled to participants. We ensured that the first trial following a rule change featured a word that would elicit an error under the response contingencies of the previous rule. This task design allowed us to probe memory for words that occurred around rule changes, operationalized here as event boundaries. Further, it enabled us to use computational models to assess the role of prediction error in shaping memory representations.

Before beginning the task, participants were informed about the set of four possible hidden rules: whether an item is (1) smaller than a backpack, (2) larger than a backpack, (3) natural in origin, or (4) manmade. They were instructed that on each trial, one of these four hidden rules was "active", and that they would sometimes change without warning (Fig. 1). The rule changes in the task were pseudorandomly preallocated across the 4 runs to guarantee an equal number of trials per active rule across the experiment. We also ensured that an equal number of rule transitions would be experienced in each run (8) with a set number of items in each run (56). Unsignaled rule shifts happened every 6–8 trials, jittered across the experiment. We reasoned that, due to the deterministic and binary nature of the rules, participants would be able to infer the correct rule by the end of the event.

Free recall task

At the end of each run, participants performed a typed free recall task[29]. We designed the task to be similar to spoken recall by enforcing entry of one word at a time and disallowing edits. Specifically, whenever a participant finished typing a word and hit the spacebar, the word disappeared. This also removed participants' ability to use previously typed words as cues. Due to minor spelling errors, we implemented a "spell-check" algorithm to correct typos (see Supplemental Methods: Spell check). Participants were given a minimum of 3 min to recall as many words as possible before a button appeared enabling them to move onto the next run.

Math distractor task

To ameliorate confounds associated with the rehearsal of words in working memory, we included a distractor task between the WRIT and the recall task. In this task, participants judged whether a given equation was correct or incorrect and received binary feedback. Equations were generated to be of the form $A + B + C = D$, where A, B, and C were single-digit integers[30]. The equations were made incorrect or correct by adding 1 to the D term 50% of the time. New equations were continually presented until the 10-second distractor interval ended.

Participants

We recruited 95 healthy younger adults (age range = 18–36 years old) for this study using the Prolific research platform. Of these, 13 were removed from analyses due to below-chance performance on the WRIT (mean accuracy less than 0.5). Another 14 were removed due to not adequately learning the rules for enough blocks in the task. Finally, 2 participants were removed for not performing one of the recall runs. Our final sample was 66 participants (mean age = 30.07 years, 32 female participants, 34 male participants, sex identified using self-report). All participants were compensated $10 for performing the study. All participants gave informed consent, and procedures were approved by the Washington University in St. Louis Institutional Review Board.

Reinforcement learning model

Reinforcement learning provides a successful framework for understanding how people use reward learning to choose between options[31]. In this framework, participants are modeled as tracking the relative value of the four rules and using that to guide their decisions[22]. For the current study, we developed a reinforcement learning model that learns a set of weights of the 4 rules, W, and then uses this to decide whether to respond 'yes' or 'no' to a stimulus. The value of saying 'yes' is calculated as the weighted sum of features present in the stimulus:

$$V(yes) = \sum_{f \in S} W(f) \tag{1}$$

Here, the model calculates the value (V) of the 'yes' response as the sum of the stored weights for each feature present in the stimulus (S). The stimulus on each trial is modeled as a binary feature vector (f), where for each of the two stimulus dimensions (size and origin) the stimulus can have a value of 0 or 1. If items were "small", they had a 0 in the first dimension, and if they were "large", this was a 1. Meanwhile, "natural" and "manmade" corresponded to 0 and 1 in the second dimension, respectively. Therefore, an item like a baseball, which is small and manmade, would have the feature vector [0, 1]. The value of 'no' is the sum of stored weights for the features that are not present in S. Upon receiving feedback, the model updates each weight using a Rescorla-Wagner update rule:

$$W^{new}(f) = W^{old}(f) + \eta\left[R_t - V(\text{choice})\right] \forall f \in \text{choice} \tag{2}$$

Here $\left[R_t - V(\text{choice})\right]$ is the reward prediction error (RPE) between the value of the choice ('yes' or 'no') and the reward outcome R_t. This prediction is multiplied by the learning rate η, a free parameter, to update all the feature weights that were present in the stimulus if the choice was 'yes', or the weights that were not in the stimulus if the choice was 'no'. We update the value of all the feature weights associated with the stimulus or its inverse as we cannot be sure of the exact feature the participant is responding to. The weights of the non-chosen features are decayed to 0 according to a decay parameter d, another free parameter[22,32].

$$W^{new}(f) = d \times W^{old}(f) \forall f \notin \text{choice} \tag{3}$$

Finally, the model uses the value of the stimulus to select whether to accept or reject the stimulus, that is, to respond yes or no. For instance, if the weights for small and manmade are high, the model would be more likely to respond "yes" to "baseball". This choice is made using the soft-max choice rule:

$$p(yes) = \frac{e^{\beta V(\text{yes})}}{e^{\beta V(\text{yes})} + e^{\beta V(\text{no})}} \tag{4}$$

Here β is an inverse-temperature parameter that controls the degree to which the agent will explore or exploit its current representation. When β values are high, behavior is more exploitative, and for low β values it is more exploratory.

Thus, the model had 3 free parameters (η, β, d), which we fit using a maximum a-posteriori approach for each participant separately, with priors set based on previous work[33]. A parameter recovery analysis (see Supplement: Parameter Recovery) indicated that this model was robustly identifiable from behavior.

In addition to examining the relationship between behavior and reward prediction errors from this model, we examined its relationship with the "certainty" of the weights over the rules. First, we converted the weight matrix (W) to a set of probabilities using a logistic transformation:

$$p(rule) = \frac{e^{(W_{rule} - \max(W))}}{\sum_{i \in \text{rules}} e^{(W_i - \max(W))}} \tag{5}$$

This produces a probability distribution of rules for each trial in the task. If the probability distribution is uniform, uncertainty about the current is maximally large. Therefore, deviances from a uniform distribution, measured using Kullback-Leiber (KL) divergence[34] indicate increased rule certainty.

Word rule inference task analyses

Determining subjective boundary points. As the sequence of items and order of rule events was assigned randomly for each participant, we marked event boundaries for each participant individually. We reasoned that, in the WRIT, people would use the first error after a rule switch as a signal that they needed to find the new rule. Typically, this first error aligned with the first trial of the event because it was an incongruent item by design. However, this was not always the case. Discrepancies could arise due to participants not finding the previous rule or making an incorrect choice by accident. Therefore, we defined "subjective boundary points" for each participant as the first trial after each rule shift on which they received zero points.

Overall, there was strong alignment between the subjective boundary point and the first trial of the event ($\mu = 82.1\%, \sigma = 10.5\%$). In analyses relative to the subjective boundary, the trial at $t + 1$ after the boundary is called the "post-boundary" trial, $t - 1$ is called the "pre-boundary", and all other trials are marked as "non-boundary".

Performance. To assess performance improvements over the course of the task, we ran a linear mixed-effects model of the form:

$$\text{reward} \sim \text{position\_relative\_to\_boundary} \times \text{run} + (1|\text{participant}) \tag{6}$$

For the run variable, the first run was modeled as 0 and each further run was incremented by 1. This enabled us to examine the degree to which participants improved at the primary task across runs. The variable indicating the position relative to boundary represented the number of trials between a given word and the nearest boundary. It was positive if it was preceded by the nearest boundary and negative if it occurred before the nearest boundary.

Response time and post-error slowing

As participants adapted to a new hidden rule over the course of an event, we expected response times to decrease. Along with this decrease, we expected to find post-error response times to be elevated as a measure of inference. We reasoned that trials following no reward would be slower than trials following reward, indicating post-error slowing[35]. Finally, we wanted to account for boundary-related slowing over and above these more general post-error slowing and event-related speedup effects. To examine these potential effects on response times, we ran a hierarchical mixed-effects model of the following form:

$$\log(\text{RT}) \sim \text{trial\_within\_event} + \text{previous\_no\_reward} + \text{boundary\_label} + (1|\text{participant}) + (1|\text{word}) \tag{7}$$

Here, the first trial within an objective event (defined by a rule shift) is modeled as 0 and each trial afterwards is incremented by 1. The variable indicating whether a previous trial provided reward was coded as 0 if the previous trial was correct and 1 if the previous trial was incorrect such that a positive effect indicates post-error slowing. Finally, the boundary label variable is a categorical regressor, where the item's position is coded such that the trial at $t+1$ after the boundary is called the "post-boundary" trial, $t-1$ is called the "pre-boundary", and all other trials are marked as "non-boundary". Non-boundary[4] is modeled as the base factor. We log-transform the RT data according to recommendations by Lo and Andrews[36].

Free recall analyses

Basic free recall analysis. In line with other work analyzing free recall, we sought to measure the existence of primacy, recency, and temporal contiguity effects in our free recall (see Kahana, Diamond, and Aka, 2022[37] for review). We measured primacy and recency using a serial position curve predicting the probability of recalling words from their serial position in WRIT. High recall of words from early positions would be evidence of primacy. High recall of words from late positions would be evidence of recency. In addition, data distribution was assumed to be normal, but this was not formally tested.

We measured temporal contiguity using a lag conditional response probability approach[38]. This is done by measuring the lag between a recalled word i and the next recalled word j and looking at the probability of that lag based on the other available lags. Both the serial position curve and conditional response probability analyses were conducted using the psifr package[39].

Analyses of events structuring recall

Transitions between events. Recent evidence points to event boundaries as "anchor points" in recall. When transitioning between events, participants are more likely to jump to boundaries than information that occurred in the middle of an event[40]. To assess this, we calculated the likelihood of transitioning to specific positions when moving to a new event. For each participant, we took the position of items that were recalled after a shift in event relative to an event boundary. We used this approach to identify all the instances during recall wherein two consecutive words belonged to different events. For example, if a participant's first recalled item was from event 2 and they next recalled an item from event 4, we examined the relative location of the event 4 item with respect to its nearest boundary during encoding. This provided us with the relative proportions with which each participant used boundary items as anchor points.

Adaptive ratio of clustering. To assess the degree to which events and item rule categories served as features by which participants clustered their recall, we used the adjusted ratio of clustering (ARC) score proposed by Roenker, Thompson, and Brown (1971)[41]. This score is calculated such that it considers the total number of category repetitions observed in the subject's recall (R_{obs}). The maximum number of possible category repetitions, designated $R_{\max}$, is given by the following:

$$R_{\max} = N - k \tag{8}$$

Here N is the total number of items recalled and k is the number of categories presented. Finally, the chance number of category repetitions, $E(R)$ is calculated as follows:

$$E(R) = \left(\frac{1}{n}\sum_i n_i^2\right) - 1 \tag{9}$$

Where n_i is the number of items recalled from category i. The final formula to calculate the score is thus:

$$ARC = \frac{R_{\text{obs}} - E(R)}{R_{\max} - E(R)} \tag{10}$$

Spell check. Because we asked people to recall words by typing them using their keyboard, our recall analyses needed to account for any misspelled words. In order to recover these mistyped words, we developed a 'spell check' algorithm that maps common typos to an item from the word pool. This algorithm first tags words that were not present in the list of words a participant saw at encoding. Next, the Levenshtein edit-distance is calculated between these words and each of the encoded words, with a substitution cost set to 2. If the edit distance between a potentially misspelled word and a single encoded word is less than or equal to 2, it is deemed to be a typo. For instance, if a person saw 'daisy' and wrote 'daist', that would be corrected as a typo. If there were more words at encoding that fit within the edit distance, the misspelled word would be left uncorrected. For instance, if a participant saw "cat" and "car" but wrote "cay" it would be left as "cay" because it is unclear whether the participant had misspelled "cat" or "car".

Hierarchical mixed effects models predicting memory. After fitting the reinforcement-learning model, we used the best-fitting parameters for each subject to extract reward prediction error (RPE) values and rule certainty for each trial of each participant. We then decomposed the RPE into its magnitude and sign and used these as separate regressors in a hierarchical mixed effects model:

$$\text{recall success} \sim \text{RPE}_{\text{abs}} \times \text{RPE}_{\text{sign}} + (1|\text{participant}) + (1|\text{word}) \tag{11}$$

Our model included a random intercept for each participant and for each word. The model used a binomial link function to predict the binary outcome of recall success (1 if a word was later recalled, 0 if not). The regressor for the value of the RPE was mean-centered.

We also fit a hierarchical mixed effects model with certainty as a predictor of recall success:

$$\text{recall success} \sim \text{certainty} + (1|\text{participant}) + (1|\text{word}) \tag{12}$$

This model again included a random intercept for each participant and each word and the regressor for the certainty was mean centered.

Finally, we fit a combined model including both certainty and RPE predictions of recall success:

$$\begin{aligned}\text{recall success} \sim \text{RPE}_{\text{abs}} \times \text{RPE}_{\text{sign}} + \text{certainty} \\ + (1|\text{participant}) + (1|\text{word})\end{aligned} \tag{13}$$

Preregistration. Neither the study reported in the primary manuscript, nor the replication reported in the supplement were preregistered.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Results

We designed a task that required participants to uncover a hidden rule by making judgments about individual words. The hidden rule underwent unsignaled changes, which required participants to rapidly adapt their model of the ongoing event. For example, if a participant thought the active rule was "natural" and they saw the word "daisy" they would respond "yes" (Fig. 1). However, if the hidden rule had changed to "manmade", they would receive 0 points. This would indicate that the rule was no longer "natural", requiring participants to infer this new hidden rule. We operationalized these rule shifts as event boundaries, and subsequently tested free recall of all words in the task to investigate the role of event boundaries and rule uncertainty in shaping the structure of episodic memory.

WRIT performance reveals event boundaries at rule transitions

Performance on the WRIT, measured in both accuracy and RT, showed evidence that people treated rule shifts as event boundaries (Fig. 2). A linear

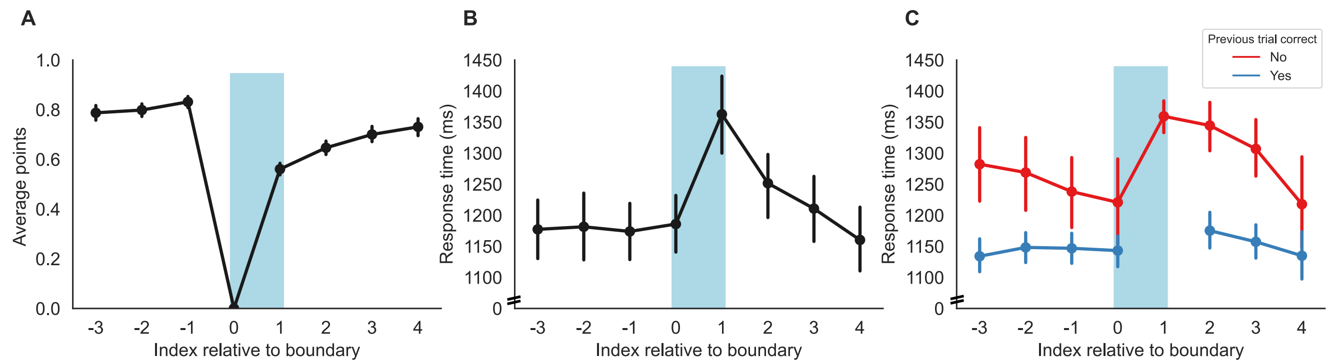

Fig. 2 | Rule shifts reliably induce event boundaries. A Task performance was high before a rule shift, zero by definition at the boundary, and then increased over the next few trials as participants discovered the active rule. **B** Participants were slowest at the post-boundary position, as this is the first trial after it is clear that the hidden rule has changed. Responses became faster as participants discovered the new hidden rule. **C** Response times as a function of whether the previous trial was rewarded (blue) or not (red). Post-error slowing is observed across all trials, with a sharp increase on trials occurring after the incorrect response to the boundary trial. Together, these behavioral patterns suggest that participants perceived rule shifts as event boundaries in the task. Error bars represents 95% confidence intervals ($n = 66$). The shaded blue box indicates the boundary area.

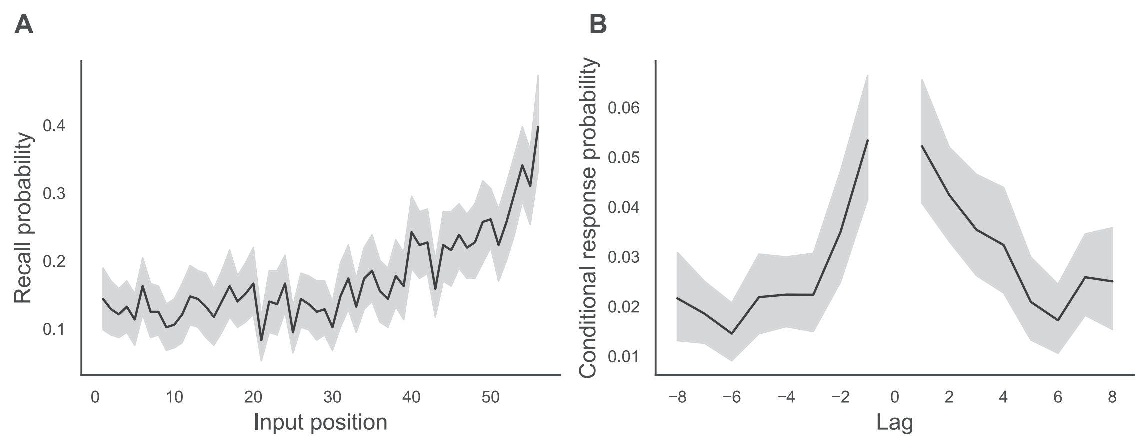

Fig. 3 | General recall characteristics. A Serial position curve depicting the probability of a work being recalled as a function of its serial position. There was a recency effect where the final positions are better recalled than those in the middle, but no primacy effect. **B** Conditional response probability curve showing evidence of canonical temporal contiguity effects in recall. The lags have been truncated to display only −8 to +8 to focus on the temporal contiguity effect (full version in Supplementary Fig. 1). Shading is 95% confidence interval ($n = 66$).

mixed effects model showed that response accuracy decreased following a rule shift compared to the stable performance they achieved before, and then increased across the event ($\beta_{position\_relative\_to\_bound} = -0.115, p < 0.001, CI95\% = [-0.15, -0.08]$) (Fig. 2A). This effect is negative as the items prior to the boundary have a negative coding and the items after a boundary have a positive position. Participant performance also improved across runs ($\beta_{run} = 0.065,\ p < 0.001, CI95\% = [0.03, 0.10]$). We also observed an interaction between run and position within the event ($\beta_{position\_relative\_to\_bound\ X\ run} = 0.022, p < 0.001, CI95\% = [0.01, 0.03]$). These results suggest that the error caused by the rule shift required participants to update their working event model.

Another common behavioral indicator of event segmentation is post-boundary slowing[2,12,42]. However, we note that post-boundary slowing could be linked to task-switch costs[43]. Consistent with this phenomenon, a linear mixed effects model showed that participants slowed down after the rule shift ($\beta_{post\_boundary} = 0.038, p = 0.004, CI95\% = [0.012, 0.064]$; Fig. 2B). Importantly, this model accounted for post-error slowing ($\beta_{prev\_no\_reward} = 76.70, p = 0.007, CI95\% = [21.28,\ 132.13]$) and general speeding over the course of an event ($\beta_{trial\_within\_event} = -0.011, p < 0.001, CI95\% = [-0.016, -0.006]$). Therefore, the slowing after a rule shift can be uniquely attributed to the participants finding out that the active rule has changed. We also found a general speedup for the boundary trial in this model ($\beta_{boundary} = -0.058, p < 0.001, CI95\% = [-0.085, -0.032]$). There was no difference in speed for pre-boundary and non-boundary items when accounting for general post-error slowing and within-event speedup ($\beta_{pre\_boundary} = 0.003, p = 0.814, CI95\% = [-0.020, 0.026]$).

Taken together, these results suggest that errors following stable performance in an event were treated differently from errors elsewhere in the task. This pattern of results is highly consistent with rule transitions inducing event boundaries during task performance.

Overall recall performance

Because participants had to recall words encountered in a demanding task, we wanted to ensure that free recall performance did not deviate strongly from expectations set by prior studies. On average, participants recalled 17.4% of the words they encountered ($\mu = 0.174, \sigma = 0.065$). While this percentage is low compared to previous free recall studies, the length of the recall list length (56) is much longer, so a smaller percentage still reflects a sizable number of words recalled[12,44]. We observed strong evidence for a recency effect (Fig. 3A), but interestingly, no evidence for a primacy effect. The lack of primacy effect may be attributed to the immediate requirement for participants to perform rule inference, reducing the resources that can be spared for encoding[44]. Recall performance also featured a high degree of temporal contiguity (Fig. 3B), with participants tending to recall sequences of words they experienced close in time in the WRIT[37]. However, we did not find evidence for the standard forward asymmetry in recall (mean conditional recall probability of lag 1 = 0.05, mean conditional recall probability of lag -1 = 0.533, $t(65) = -0.14, p = 0.890$, $Cohen's\ d = 0.02$, CI 95%=[-0.37,0.32] for lag 1 vs lag -1).

Impaired recall of post-boundary items

Having validated both performance in the WRIT and the recall task, we turned our attention to the effects of event boundaries on retrieval. Interestingly, we found no evidence for boundary-related memory enhancements predicted by extant theories such as EST[14]. Boundary items showed an inconclusive difference in recall from non-boundary items (Mean boundary recall = 0.16, mean non-boundary recall = 0.17, $t(65) = -1.65, p = 0.103, Cohen's\ d = -0.19, CI95\% = [-0.53, 0.16]$). However, pre-boundary items were recalled more often than non-boundary (Mean pre-boundary recall = 0.19, $t(65) = 2.22, p = 0.030, Cohen's\ d = 0.25, CI95\% = [-0.1, 0.59]$). While lack of a boundary-specific enhancement has been reported before[12], we found evidence that event boundaries impaired memory. Boundary items were recalled less often than pre-boundary items ($t(65) = -2.89, p = 0.005$, $Cohen's\ d = -0.39$, $CI95\% = [-0.74, -0.04]$). Most importantly, post-boundary items were recalled less often than items in all other positions (mean recall of post-boundary = 0.13, post-boundary against non-boundary $t(65) = -4.29, p < 0.001, Cohen's\ d = -0.51, CI95\% = [-0.86, -0.16]$, post-boundary against pre-boundary $t(65) = -4.96, p < 0.001$, $Cohen's\ d = -0.67, CI95\% = [-1.03, -0.32]$ and post-boundary against boundary $t(65) = -2.11, p = 0.038, Cohen's\ d = -0.27, CI95\% = [-0.62, 0.07]$; Fig. 4).

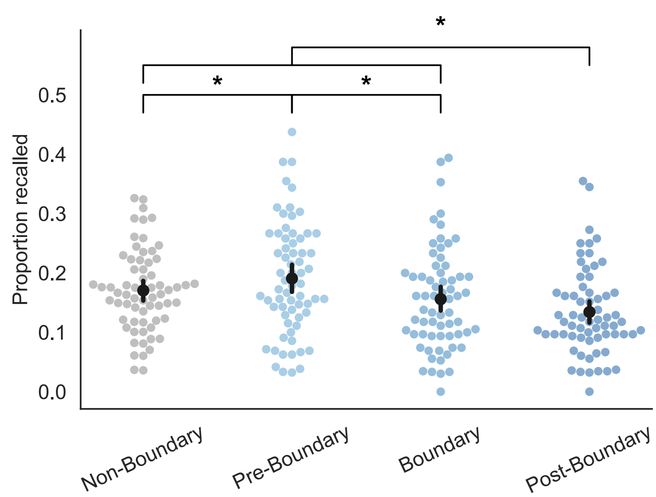

Fig. 4 | Post-boundary inference leads to a recall deficit. Words in the post-boundary position were remembered worse than all other positions. Interestingly, we found no significant effect of a recall benefit for boundary items. We also find no difference between non-boundary, pre-boundary, or boundary items. Error bars represent 95% confidence intervals ($n = 66$).

Boundaries and events structure recall

Events scaffold the structure of free recall. We next tested the extent to which event structure influenced free recall. In line with previous work[40], we reasoned that participants would display temporal contiguity effects at the level of events. To this end, we calculated the conditional response probability based not on serial position, but rather on which event an item belonged to (Fig. 5A). Unlike a standard conditional response probability, where self-same transitions indicate item repetitions, here they indicate the probability of staying within an event during recall. We found that participants tended to cluster their recall with words that occurred in the same event (mean probability = 0.037). We found no evidence for forward asymmetry for the recall of next event items at lag 1, though this could be due to a lack of power to test this specific comparison (mean of lag 1 = 0.03, mean of lag -1 = 0.02, test of lag 1 vs lag -1 $t(65) = 1.52, p = 0.134, Cohen's\ d = 0.27, CI95\% = [-0.07, 0.62]$). However, when examining all forward vs backward across-event transitions, we found that transitions tended to occur in the forward order (mean of positive lags = 0.13, mean of negative lags = 0.02, test of negative lags vs positive ones $t(65) = 7.1, p<0.001, Cohen's\ d = 1.28$, $CI95\% = [0.90, 1.66]$). In sum, these analyses revealed a tendency to stay within an event during recall, and that transitions across events tended to occur in the forward direction.

Event boundaries anchor transitions between events in free recall. Above, we found that participants tended to cluster items of the same event during recall, but that they sometimes transitioned between events. Next, we assessed the nature of these transitions. In line with previous work[40], we predicted that event-boundary timepoints would serve as anchors during free recall. That is, we predicted that if participants transitioned to recalling items from a new event, they would be more likely to transition to words presented on an event boundary words than those in the middle of events. We test this prediction of event-level transitions during recall by measuring degree to which participants recalled consecutive words from different events (See Methods: *Transitions between events*). For these analyses, we included only participants

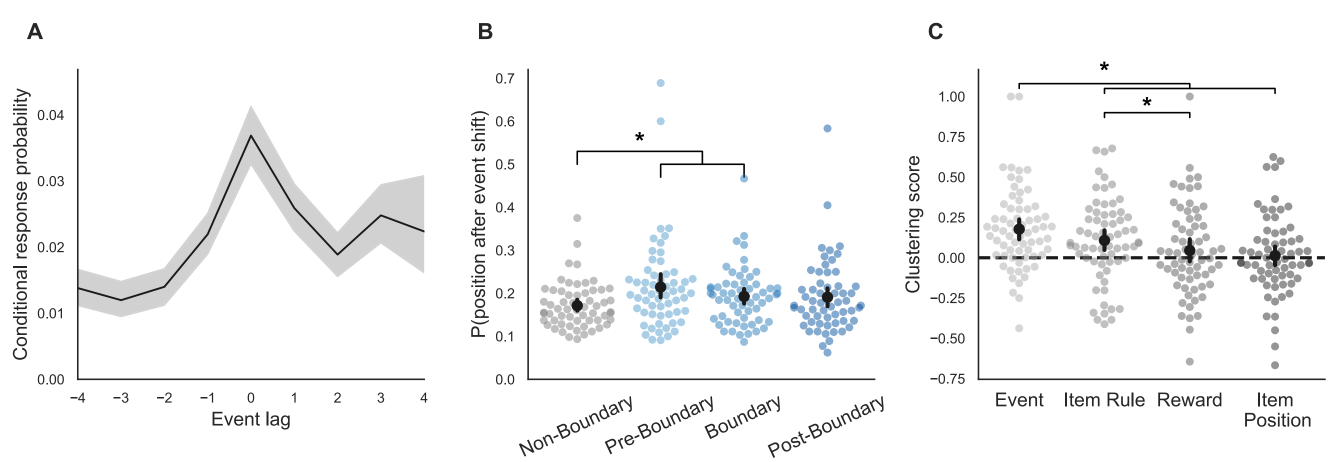

Fig. 5 | Events organize free recall. A We calculated a conditional response probability curve using the event lag to examine the degree of temporal contiguity while recalling events. Participants were most likely to recall words from the same event as the previously recalled word. **B** The probability of recalling a word at a given position after transitioning events. Interestingly, pre-boundary items were most commonly used as the index into an event. ($n = 59$) **C** Using the adjusted ratio of clustering approach we found above-chance clustering of recalled items based on the event they were experienced in, as well as the rule of the item. Error bars and shading represent 95% confidence intervals ($n = 66$).

that had each type of anchor transition (n = 59). Indeed, items at boundaries were used as anchors more often than non-boundary items (mean boundary anchoring = 0.19, mean of non-boundary anchoring = 0.17, $t(58) = 2.83, p = 0.006$, $Cohen's\ d = 0.36$, $CI\ 95\% = [-0.01, 0.73]$) (Fig. 5B). Somewhat unexpectedly, we found that pre-boundary items also anchored recall more often than non-boundary items (mean of pre-boundary = 0.21, $t(58) = 3.78$, $p < 0.001$, $Cohen's\ d = 0.52$, $CI95\% = [0.15,\ 0.89]$). We found no evidence that people differed in their tendency to use either pre-boundary or boundary items as anchors when transitioning between events ($t(58) = 1.7, p = 0.094$, $Cohen's\ d = 0.25 CI95\% = [-0.12, 0.61]$). Though pre-boundary and boundary items numerically anchored recall more so than post-boundary items, they did not differ significantly (mean of post-boundary = 0.19, test against pre-boundary: $t(58) = 1.43, p = 0.156$, $Cohen's\ d = 0.25$, $CI95\% = [-0.12,\ 0.61]$; test against boundary: $t(58) = 0.13, p = 0.89$, $Cohen's\ d = 0.02$, $CI95\% = [-0.34,\ 0.39]$). In sum, we found evidence for an increased tendency to transition across events to items that occurred at or before (but not after) an event boundary.

Recall clusters by event and by rule present during encoding. We next assessed whether, aside from the contiguity of events and event boundaries, recall clustered along other categorical dimensions. To do this, we calculated an adjusted ratio of clustering to examine the nature of recall clustering regardless of transition order. We found that participants' recall was not only clustered by the event in which that word was experienced (mean clustering by event = 0.17, $t(65) = 5.71, p < 0.001$, $Cohen's\ d = 0.70$, $CI95\% = [0.429,\ 0.978]$), but also by the hidden rule that was active when the item was presented (mean clustering by rule = 0.11, $t(65) = 3.42, p = 0.001$, $Cohen's\ d = 0.42$, $CI95\% = [0.16,\ 0.68]$; Fig. 5C). We found no evidence that participants used reward as a dimension to organize their free recall (mean clustering by reward = 0.05, $t(65) = 1.31, p = 0.193$, $Cohen's\ d = 0.16$, $CI95\% = [-0.09, 0.41]$) (though see Horwath et al. 2023[45]). Finally, we examined whether the position of an item with respect to an event was used to organize recall. For instance, we tested whether pre-boundary items were often followed by other pre-boundary items but found no evidence for such organization (mean clustering by item position = 0.01, $t(65) = 0.423, p = 0.673$, $Cohen's\ d = 0.05$, $CI95\% = [-0.19,\ 0.30]$).

A reinforcement learning model predicts participants' recall. Next, we investigated how the reward prediction error and rule certainty measures from the reinforcement-learning model related to recall success. Even though we ran two separate models predicting recall success as a function of these components individually, we found that the combined model significantly outperformed both (combined vs certainty only $\chi2(3)$ =15.75, p=0.001, combined vs RPE only $\chi2(1) = 19.00, p < 0.001$). For this reason, we focus on the results of the combined model.

First, following previous work, we predicted that the absolute magnitude of RPEs would be positively associated with increased memory performance[32,46]. Counterintuitively we find a negative effect of absolute RPE on memory ($\beta_{RPE_{abs}} = -0.162, p = 0.013$, $CI95\% = [-0.291, -0.034]$). We find no main effect of value ($\beta_{RPE_{val}} = -0.040, p = 0.091$, $CI95\% = [-0.086,\ 0.006]$). We do find a marginally significant interaction between the value and sign of RPE ($\beta_{RPE_{abs}:RPE_{val}} = -0.123, p = 0.065$, $CI95\% = [-0.253, 0.007]$), this effect is one of the primary differences between the replication and the primary sample (**Supplement: Replication**). Upon further examining the RPE as a function of trials (Fig. 6A, B), one can see that in most trials there is a small positive prediction error. This could be somewhat of an artifact of how the weight decay works in the model. On any given trial only one rule is active but values of two rules are always updated based on the choice and reward. This leads to a situation where the non-relevant feature alternates it constantly generates a small positive RPE. Interestingly, the model without this decay fits worse across all participants (See **Supplement: Model Comparison**).

To account for this, we developed a certainty measure using the KL divergence of the weight matrix from a uniform distribution (**Methods: reinforcement learning model**). As the model learned more asymmetric weights with a peak on the relevant dimension, the alternation of the weights on the irrelevant dimension that produces the positive RPE would still result in the same certainty measure. Upon fitting the combined model we found certainty was a strong predictor of memory, such that higher certainty about the rule distribution was linked to higher recall performance ($\beta_{certainty} = 5.469, p < 0.001$, $CI95\% = [3.047, 7.892]$) (Fig. 6C, D).

Discussion

Everyday life is full of uncertainty, requiring interaction and inference to discern important information in the environment. However, with experience, we learn the structure of the world sufficiently well to predict how things typically unfold. EST argues that we do so by forming event models that predict perceptual input[1]. However, much of the prior literature has studied passive viewing of events as they unfold, and has often featured paradigms where uncertainty is difficult to quantify. Other research has studied how people interactively resolve uncertainty, but these experiments offer little insight into the influence of event structure on this process. Here, we examined the formation of event models in inferential, interactive events. Contrary to predictions from EST, we observed memory *deficits* for information encountered after event boundaries, and no evidence for memory enhancements for information around event boundaries. This suggests that constructing event models in environments where one can interactively reduce uncertainty is not conducive to encoding episodic memories. At the same time, we found that event structure guided free recall, such that participants used events as a fundamental organizing property. Finally, in line with previous work on inferential hypothesis testing, behavior in the WRIT was well described by a reinforcement-learning model[22,47]. However, in contrast to previous studies[32,46], we found a negative relationship between memory and the magnitude of the RPE. Only smaller RPEs were associated with greater recall success.

The influence of event boundaries on episodic memory has been the subject of much investigation[3,12,48]. In recent work, Richmond and Zacks posit that event segmentation and event model construction processes result in increased encoding for items immediately surrounding an event boundary[14] (see also Clewett et al., 2019[15]). In this study, however, we do not see evidence for a consistent recall benefit for pre-boundary or boundary items. Importantly, we found evidence in conflict with a prediction of EST: recall was systematically worse for items encoded following an event boundary. This may suggest that event model construction is harmful to rather than helpful for encoding episodic memory under some circumstances. What could account for this unexpected result?

One intriguing possibility is that this reduction in memory is driven by the effort requirements of active inference. Indeed, research on cognitive control suggests that the exertion of mental effort carries an intrinsic cost[49]. We believe that the process of inferring the hidden rule in the WRIT requires many control-demanding computations (e.g., manipulation of information in working memory, task switching costs[50]). The cost associated with these computations may reduce the availability of attentional resources for encoding information in long-term memory. Another possibility is that the post-boundary memory deficit is driven by the lack of a representational scaffold immediately after rule shifts. Future investigations may disambiguate between these hypotheses. Altogether, our findings reveal a complex interplay between event segmentation and interactive, inferential processes in deciding the fate of episodic memories. Importantly, due to the design of many prior studies, this interplay has until now gone undetected.

Free recall is well known to be affected by the structure of encoded information. For example, word recall order is strongly shaped by temporal context (Howard and Kahana 2002) and semantic category[51]. Here, we found that higher-order event structure, delineated by shifting rules, served

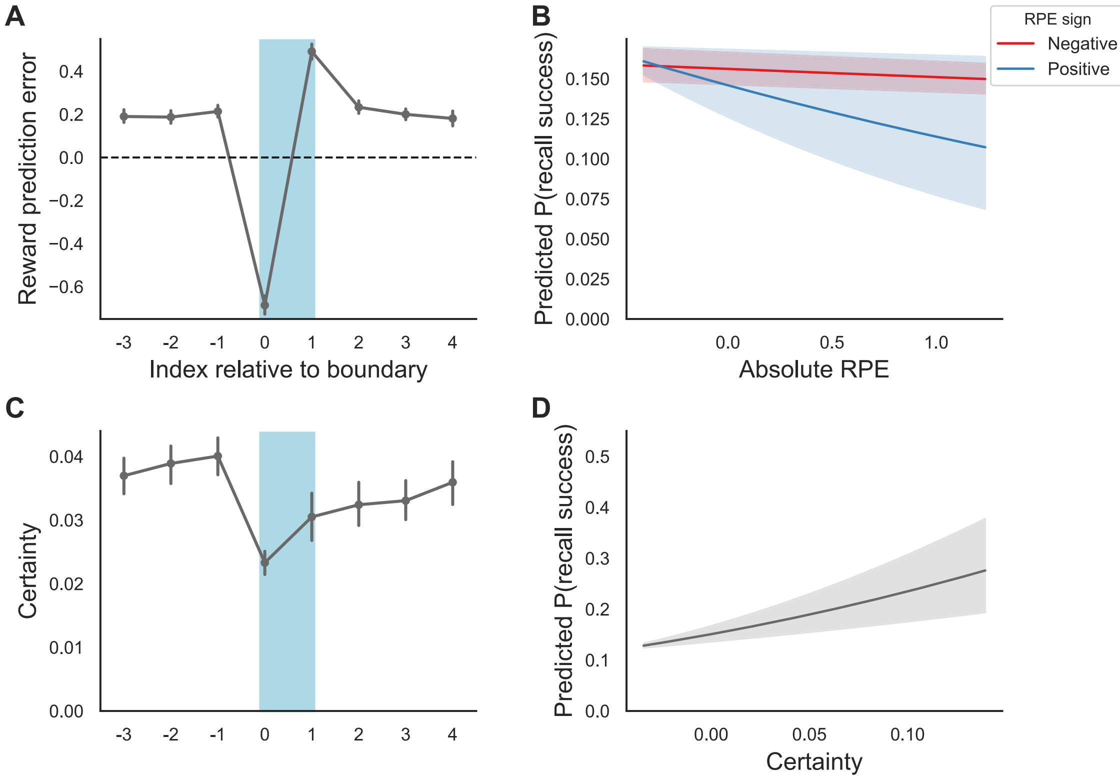

Fig. 6 | A reinforcement learning model predicts recall. **A** Across the task, participants generally experience small positive RPEs. This happens because performance improved across the duration of an event. Moreover, each trial provides RPEs for both a relevant and irrelevant stimulus dimension. As the irrelevant feature alternates participants will experience a small positive RPE. The event boundary time point has a large negative RPE as the weights for a specific rule go from being consistently rewarded to unrewarded when the rule has changed. **B** The predicted probability of recall success changes as a function of the magnitude of RPEs. Positive RPEs are negatively predictive of recall success such that larger positive RPEs are coupled with lower odds of recalling an item. Negative RPEs meanwhile, while qualitatively flatter, do not statistically differ in their effect on recall. **C** Across the task, participants gain certainty about the active rule before becoming uncertain at event boundaries, then must slowly build up certainty once again. **D** Increased certainty is coupled with a higher predicted probability of recall. (Error bars are 95% confidence interval, shading is 95% confidence interval of the HLM effect estimate (n = 66))

as a scaffold for organizing recall. Specifically, participants anchored their recall to items that served as event boundaries, as well as pre-boundary items. This finding aligns with a recent study by Michelmann and colleagues (2023), where participants were found to jump between event boundaries when remembering events in a movie. However, our paradigm features several important differences from this study. In a Hollywood-style film, event transitions are both passively viewed and purposefully telegraphed to an observer. In the WRIT, however, participants interactively discover when event transitions occur. In particular, the boundaries in our task were likely perceived *after* encoding the boundary item, when participants received feedback. Our results suggest that the way events structure recall depends on whether their boundaries can be either simply observed and instantaneously processed, or if they need to be actively inferred from interactions with one's environment.

Previous work incorporating reward into event cognition has found that unsignaled surprise heightens memory and forms event boundaries[32,52]. Here, we found the inverse, that large magnitude RPEs dampen memory. This may be due to the structure of our task, where the largest magnitude negative RPEs occur at the event boundary by design, and the largest positive RPEs occur directly afterwards. Though it should be noted that even though we replicated the effect of higher magnitude positive RPEs predicting memory decrements in a follow-up replication, we found that the negative RPEs showed a different trend (Supplemental Results: Replication). Nonetheless, this reveals that memory is not unilaterally affected by surprise, a distinction not made by theories of event cognition.

Theories of latent cause inference predict increased uncertainty after rule shifts, at the same time, our reinforcement learning model encodes particularly strong negative RPEs. Indeed, the RPEs signal the need for a revaluation of the internal task representation. This suggests that our model, even though not explicitly formulated in terms of latent cause inference, accounts for uncertainty reduction over latent causes. Consistent with this, KL divergence (a measure of uncertainty over rules) strongly predicted memory performance, in line with predictions from latent cause inference-based theories of event cognition[16]. Future experiments could aim to better disentangle latent cause inference and reinforcement learning to enable different modeling approaches, as previous work has highlighted the strong connections between latent cause inference and event cognition[17–21].

Limitations

The goal of the present study was to develop an interactive approach to studying the role of event structure in shaping episodic memory. In our paradigm, like in many events we encounter in daily life, people must interact with their environment to understand which set of behaviors fit a given situation. A key limitation is that the task we used to study this was considerably less naturalistic than studies of event memory that involve movie viewing[3,9,40]. However, the use of word stimuli and the use of free recall as the key measure enabled us to manipulate the nature of event boundaries and to probe specific content and structural components of recall. Using individual words as trials, we were able to define transitional moments and track performance and uncertainty reduction across distinct epochs. This enabled us to discover the effects of event boundaries on memory that run counter to predictions of EST. Furthermore, the gain in information from individual trials is slower and more easily modeled than in more naturalistic stimuli. This allowed us to better understand how uncertainty reduction interacts with event model construction, as well as better investigate reaction times, as in the task-switching literature[43]. Our research may inspire new studies that use more naturalistic paradigms to study the role of inference in event cognition.

One limitation of all free recall experiments is that participants tend to recall only a fraction of the studied list. However, future studies can mitigate this by using a recognition memory paradigm, which would enable one to assess memory for each item.

A third limitation of the current work is that event boundaries are being inferred from changes in response time and task behavior rather than directly indexed. In conjunction with the simplicity of the task, this might raise a question of whether "event boundaries" per se are being induced. However, this experimental logic has been applied to many contemporary studies over similar topics[2,12] and the behavioral effects we observe (e.g., slowed RTs after a shift) are consistent with this operationalization of event boundaries. Though previous studies with similar paradigms have used a reinforcement learning formalization to characterize behavior[22], it is not clear whether the neural underpinnings of our results are necessarily performed in systems thought to be connected to reinforcement learning (e.g., ventral striatum). Indeed, the reward prediction error we generate with our model could instead be proposed to be a domain-general prediction error (though we retain the reinforcement learning terminology for consistency with previous literature[22]).

A final limitation of the current work is that the experiment does not contain causal connections between events. In daily life, there are many causal reasons as to why a given scenario follows another, and there are frameworks of event cognition that put more weight on this[4,6,16]. Indeed, this could be an important distinction between the phenomena we observe and those best characterized by EST. Future work can expand this experimental paradigm by embedding causal connections between the rule shifts to better compare the predictions of these frameworks.

Conclusion

In sum, our findings reveal that the structure of events during encoding scaffolds later recall of individual items, with event boundaries serving as anchor points. Further, we found that, in a dynamic and interactive task, event boundaries inhibit rather than enhance memory for post-boundary items. This runs contrary to predictions of EST, as well as several empirical findings from studies using passive tasks. Finally, we found reduced memory for items that were experienced with larger prediction errors, particularly those with positive prediction errors. Overall, our study suggests that event segmentation and its effects on long-term memory seem to be fundamentally different in situations where participants can interactively decrease their uncertainty instead of passively waiting for perceptual input. These results deepen our understanding of the way event structure scaffolds episodic memories and can guide the development of theoretical and computational models of event cognition.

Data availability

The raw datasets for the primary study and the replication are available at the following OSF: https://osf.io/3d746/ The processed dataset for the primary study and replication are available on GitHub at the following link: https://doi.org/10.5281/zenodo.12699873.

Code availability

The code for running the experiment (jsPsych) as well as analyzing the data (custom python scripts) for the study is available on GitHub at the following link: https://doi.org/10.5281/zenodo.12699873.

Received: 23 August 2024; Accepted: 28 March 2025;
Published online: 26 April 2025

References

1. Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S. & Reynolds, J. R. Event perception: a mind-brain perspective. *Psychol. Bull.* **133**, 273–293 (2007).
2. DuBrow, S. & Davachi, L. Temporal binding within and across events. *Neurobiol. Learn. Mem.* **134**, 107–114 (2016).
3. Swallow, KhenaM., Jeffrey, M. Zacks & Richard, A. Abrams Event boundaries in perception affect memory encoding and updating. *J. Exp. Psychol.: Gen.* **138**, 236–257 (2009).
4. Radvansky, G. A. Across the event horizon. *Curr. Direct. Psychol. Sci.* **21**, 269–272 (2012).
5. Zacks, J. M. Event perception and memory. *Annu. Rev. Psychol.* **71**, 165–191 (2020).
6. Zwaan, R. A., Langston, M. C. & Graesser, A. C. The construction of situation models in narrative comprehension: An Event-Indexing Model. *Psychol. Sci.* **6**, 292–297 (1995).
7. Sargent, J. Q. et al. Event segmentation ability uniquely predicts event memory. *Cognition* **129**, 241–255 (2013).
8. Bailey, H. R., Kurby, C. A., Sargent, J. Q. & Zacks, J. M. Attentional focus affects how events are segmented and updated in narrative reading. *Mem. Cogn.* **45**, 940–955 (2017).
9. Radvansky, G. A. & Copeland, D. E. Walking through doorways causes forgetting: Situation models and experienced space. *Mem. Cogn.* **34**, 1150–1156 (2006).
10. Cowan, E. T., Chanales, A. J., Davachi, L. & Clewett, D. Goal shifts structure memories and prioritize event-defining information in memory. *J. Cognit. Neurosci*, 1–17. https://doi.org/10.1162/jocn_a_02220 (2024).
11. Wang, Y.C. & Egner, T. Switching task sets creates event boundaries in memory. *Cognition* **221**, 104992 (2022).
12. Heusser, A.C., Ezzyat, Y., Shiff, I., Davachi, L. Perceptual boundaries cause mnemonic trade-offs between local boundary processing and across-trial associative binding. *J. Exp. Psychol.: Learn., Mem., Cogn.* **44**, 1075–1090 (2018).
13. Pu, Y., Kong, X.-Z., Ranganath, C. & Melloni, L. Event boundaries shape temporal organization of memory by resetting temporal context. *Nat. Commun.* **13**, 622 (2022).
14. Richmond, L. L. & Zacks, J. M. Constructing experience: event models from perception to action. *Trends Cogn. Sci.* **21**, 962–980 (2017).
15. Clewett, D., DuBrow, S. & Davachi, L. Transcending time in the brain: How event memories are constructed from experience. *Hippocampus* **29**, 162–183 (2019).
16. Shin, Y. S. & DuBrow, S. Structuring memory through inference-based event segmentation. *Top. Cogn. Sci.* **13**, 106–127 (2021).
17. Franklin, N. T., Norman, K. A., Ranganath, C., Zacks, J. M. & Gershman, S. J. *Structured Event Memory: A Neuro-Symbolic Model of Event Cognition*. (2020).
18. Varga, N., Morton, N. & Preston, A. *Schema, Inference, and Memory*. OSF. https://doi.org/10.31234/osf.io/m9adb (2022).
19. Bein, O. & Niv, Y. Schemas, reinforcement learning and the medial prefrontal cortex. *Nat. Rev. Neurosci.* 1–17. https://doi.org/10.1038/s41583-024-00893-z (2025).
20. Lu, Q. et al. Reconciling shared versus context-specific information in a neural network model of latent causes. *Sci. Rep.* **14**, 16782 (2024).
21. Beukers, A. O. et al. *Blocked training facilitates learning of multiple schemas*. PsyArXiv. https://doi.org/10.31234/osf.io/9bptj (2023).
22. Niv, Y. et al. Reinforcement learning in multidimensional environments relies on attention mechanisms. *J. Neurosci.* **35**, 8145–8157 (2015).
23. Wilson, R. & Yael N. Inferring relevance in a changing world. *Front. Hum. Neurosci.* **5**. https://www.frontiersin.org/articles/10.3389/fnhum.Z.00189. (2012).
24. Mirea, D.-M., Shin, Y. S., DuBrow, S. & Niv, Y. The ubiquity of time in latent-cause inference. *J. Cognit. Neurosci.*, 1–13. https://doi.org/10.1162/jocn_a_02231 (2024).
25. Radulescu, A., Shin, Y. S. & Niv, Y. Human representation learning. *Annu. Rev. Neurosci.* **44**, 253–273 (2021).
26. Shen, X., Ballard, I. C., Smith, D. V. & Murty, V. P. Decision uncertainty during hypothesis testing enhances memory accuracy for incidental information. *Learn. Mem.* **29**, 93–99 (2022).

27. Barceló, F. A predictive processing account of card sorting: fast proactive and reactive frontoparietal cortical dynamics during inference and learning of perceptual categories. *J. Cogn. Neurosci*. https://doi.org/10.1162/jocn_a_01662 (2021).
28. Howard, M. W. & Kahana, M. J. A distributed representation of temporal context. *J. Math. Psychol.* **46**, 269–299 (2002).
29. Unsworth, N. & Miller, A. L. Encoding dynamics in free recall: Examining attention allocation with pupillometry. *Mem. Cogn.* **49**, 90–111 (2021).
30. Howard, M. W. & Kahana, M. J. Contextual variability and serial position effects in free recall. *J. Exp. Psychol.: Learn. Mem., Cogn.* **25**, 923–941 (1999).
31. Sutton, R. S. & Andrew G. B. *Reinforcement Learning: An Introduction*. Second edition. Adaptive Computation and Machine Learning Series. Cambridge, Massachusetts: The MIT Press. (2018).
32. Rouhani, N. & Yael N. Signed and Unsigned Reward Prediction Errors Dynamically Enhance Learning and Memory." Edited by Thorsten Kahnt and Christian Büchel. *eLife* **10**: e61077. https://doi.org/10.7554/eLife.61077. (2021).
33. Gershman, S. J. Empirical priors for reinforcement learning models. *J. Math. Psychol.* **71**, 1–6 (2016).
34. Kullback, S. & Leibler, R. A. On information and sufficiency. *Ann. Math. Stat.* **22**, 79–86 (1951).
35. Danielmeier, C. & Ullsperger, M. Post-Error Adjustments. *Front. Psychol. 2*. https://doi.org/10.3389/fpsyg.2011.00233 (2011).
36. Lo, S. & Andrews, S. To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Front. Psychol. 6*, https://doi.org/10.3389/fpsyg.2015.01171 (2015).
37. Kahana, Michael J., Nicholas B. Diamond, & Ada A. Laws of human memory. https://psyarxiv.com/aczu9/. (2022).
38. Kahana, M. J. Associative retrieval processes in free recall. *Mem. Cognition* **24**, 103–109 (1996).
39. Morton, NealW. PSIFR: Analysis and visualization of free recall data. *J. Open Source Softw.* **5**, 2669 (2020).
40. Michelmann, S., Uri H. & Kenneth A. N. Evidence that event boundaries are access points for memory retrieval. *Psychol. Sci.* 09567976221128206. https://doi.org/10.1177/09567976221128206 (2023).
41. Roenker, D.L., Thompson, C.P. & Brown, S.C. Comparison of measures for the estimation of clustering in free recall: psychological bulletin. *Psychol. Bull.* **76**, 45–48 (1971).
42. Zwaan, R. A. & Radvansky, G. A. Situation models in language comprehension and memory. *Psychol. Bull.* **123**, 162–185 (1998).
43. Rogers, R. D., & Monsell, S. Costs of a predictible switch between simple cognitive tasks. *J. Exp. Psychol. Gen*. https://doi.org/10.1037/0096-3445.124.2.207 (1995).
44. Healey, M. K. Temporal contiguity in incidentally encoded memories. *J. Mem. Lang.* **102**, 28–40 (2018).
45. Horwath, E. A., Rouhani, N., DuBrow, S. & Murty, V. P. Value restructures the organization of free recall. *Cognition* **231**, 105315 (2023).
46. Rouhani, N., Norman, K. A., Niv, Y. & Bornstein, A. M. Reward prediction errors create event boundaries in memory. *Cognition* **203**, 104269 (2020).
47. Song, M., Baah, P. A., Cai, M. B. & Niv, Y. Humans combine value learning and hypothesis testing strategically in multi-dimensional probabilistic reward learning. *PLOS Comput. Biol.* **18**, e1010699 (2022).
48. Morse, S. J., Karagoz, A., & Reagh, Z. *Event boundaries directionally influence item-level recognition memory* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/h8bj2 (2023).
49. Kool, W. & Botvinick, M. Mental labour. *Nat. Hum. Behav.* **2**, 899–908 (2018).
50. Monsell, S. Task switching. *Trends Cogn. Sci.* **7**, 134–140 (2003).
51. Polyn, S. M., Norman, K. A. & Kahana, M. J. A context maintenance and retrieval model of organizational processes in free recall. *Psychol. Rev.* **116**, 129–156 (2009).
52. Rouhani, N., Norman, K.A. & Niv, Y. Dissociable effects of surprising rewards on learning and memory. *J. Exp. Psychol.: Learn. Mem. Cogn.* **44**, 1430–1443 (2018).

Acknowledgements

We thank Emily Cowan, Mike Freund, Neal Morton, Craig Poskanzer, Angela Radulescu, and Jeff Zacks for helpful discussion. We also thank the members of the Complex Memory Lab, Control and Decision Making Lab, and Dynamic Cognition Lab at Washington University in St. Louis. A.B.K. was funded by NINDS T-32-NS115672. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

A.B.K.: Conceptualization, Methodology, Software, Analysis, Visualization, Writing – Original Draft, Writing – Reviewing and Editing. W.K.: Conceptualization, Methodology, Writing – Original Draft, Writing – Reviewing and Editing, Supervision. Z.M.R.: Conceptualization, Methodology, Writing – Original Draft, Writing – Reviewing and Editing, Supervision.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s44271-025-00243-4.

Correspondence and requests for materials should be addressed to Ata B. Karagoz.

Peer review information *Communications Psychology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: Troby Ka-Yan Lui. [A peer review file is available].