# Free recall is shaped by inference and scaffolded by event structure

Ata B. Karagoz, Wouter Kool*, Zachariah M. Reagh*

Psychological and Brain Sciences, Washington University in St. Louis

*authors contributed equally

# Abstract

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 are thought 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 is not well understood. Here, we designed a task to test how boundaries between hidden rules and uncertainty about those rules affect recall of item-level information. Participants responded to individual word stimuli, with words grouped by 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, and contrary to predictions of theories of event cognition, recall was worse for words encoded immediately after event boundaries. These findings indicate that event structure and inference play important roles in shaping episodic memory.


**Statement of relevance**

Influential theories of cognition have characterized the way people represent and remember continuous experiences as discrete events. However, these frameworks mainly assess event representations in passive viewing paradigms. Here, we developed a novel word classification task where participants' goal states created events, and sudden goal shifts caused event boundaries. Participants were next asked to recall the words with which they had interacted. This recall data revealed that event representations during encoding scaffolded recall. Critically, we also observed memory deficits for information encoded after an event boundary, which stands at odds with leading event cognition theories. While these theories predict that the process of segmenting experience into events aids memory for the items occurring close to their boundaries, we find the opposite. Our results invite changes to contemporary theories of event cognition and episodic memory, suggesting that they should account for the effects of reasoning, inference and decision making.

## Introduction

In everyday life, we are faced with a continuous stream of experience. Despite this, people tend to think about and remember continuous experiences as discrete events separated by boundaries. The field of event cognition describes this process as *event segmentation* (Zacks et al. 2007; DuBrow and Davachi 2016; Swallow, Zacks, and Abrams 2009). Event segmentation is consistent across people, even with minimal instructions, and is thought to enable us to more effectively learn from and remember the world around us (see Zacks 2019 for review). However, despite much progress in recent years, the consequences of event structure on memory representations remain unclear.

Event Segmentation Theory (EST; Zacks et al., 2007) is a major theoretical framework of how people segment events. This theory 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. EST predicts that long-term memory is scaffolded by events and influenced by boundaries between them (Zacks et al. 2007; Radvansky and Zacks 2017; Richmond and Zacks 2017). This has been supported by numerous studies showing that event boundaries lead to the binding of information within events and separation across events (Davachi and DuBrow 2016; Heusser et al. 2018; Pu et al 2022; Wang and Egner 2022).

EST also predicts that memory for information that occurs around boundaries is enhanced (Swallow, Zacks, and Abrams 2009, also see Clewett, DuBrow, and Davachi 2019 for review). In short, EST assumes that prediction errors during event boundaries establish the need for a new event model. The model facilitates this formation by opening perceptual input gates, enhancing encoding of information around the event boundary. However, very few investigations of the influence of event boundaries on item-level memory have been conducted, especially for information in long-term memory.

A present limitation of studies of event cognition is that these observations largely stem from passive viewing experiments (Swallow, Zacks, and Abrams 2009; Sargent et al. 2013; Bailey et al. 2017). However, in many real-world activities, people are not merely passive observers. Instead, they interact with their environment. For instance, merely observing people play a hand of poker might be represented and remembered differently than if you were playing the hand yourself. This reveals an important gap in our understanding of event segmentation: to what extent do theories such as EST and its body of supporting evidence extend to dynamic, active environments?

An adjacent body of research has begun to investigate how people infer latent states through interaction with the environment (Niv et al. 2015; Wilson and Niv 2012, see Radulescu, Shin and Niv 2021 for review). Here, prediction errors also play a key role, signaling the need to reevaluate hypotheses about the state of the world. This suggests an intriguing analogy between the hypotheses formed during latent state inference and the event models from 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. Confirming a hypothesis, then, can lead to the establishment of a stable event model. Thus, standard models of event cognition and more recent approaches of latent cause inference stand to inform one another. Here, we seek 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.

Previous work has shown that alternating goals can provide an organizational scaffold for memory and cause event boundaries (Polyn, Norman, and Kahana 2009; Polyn et al. 2012; Wang and Egner 2022; Cowan et al. 2024). In these studies, however, states and their shifts were cued, limiting our ability to understand the role of inference and uncertainty reduction on episodic representations. This is akin to *passively* viewing a hand of poker play out, which relies on a reactive inference process. However, many of our everyday activities involve *actively* generating and testing predictions about the state of the world.

To test how event segmentation and inference in an interactive environment affect the organization of episodic memory, we designed a novel word rule inference task (WRIT). Participants inferred an active “task rule” by interacting with the environment and receiving feedback (Wilson and Niv 2012; Shen et al. 2022; Barceló 2021). Specifically, they indicated whether a word agreed with a hidden rule, receiving rewards when they responded correctly. This hidden rule shifted at unpredictable moments, 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 (Howard and Kahana 2002).

Our findings reveal that event boundaries in interactive and inferential settings affect both the content and structure of recall. Participants showed strong evidence for the perception of event boundaries during shifts in the hidden rules. We found that participants clustered their recall based on the rule-shifting events they encountered during encoding, indicating that event structure during encoding scaffolded free recall. Critically, we also found that the probability of recalling a specific word was lowest for words immediately following an event boundary. This suggests that interactive inference processes may hinder encoding, contrary to predictions made by EST. Together, these findings indicate that actively navigating states in the world leads to structured representations that scaffold episodic memory, and inference after event boundaries may in some cases be *detrimental* to encoding.

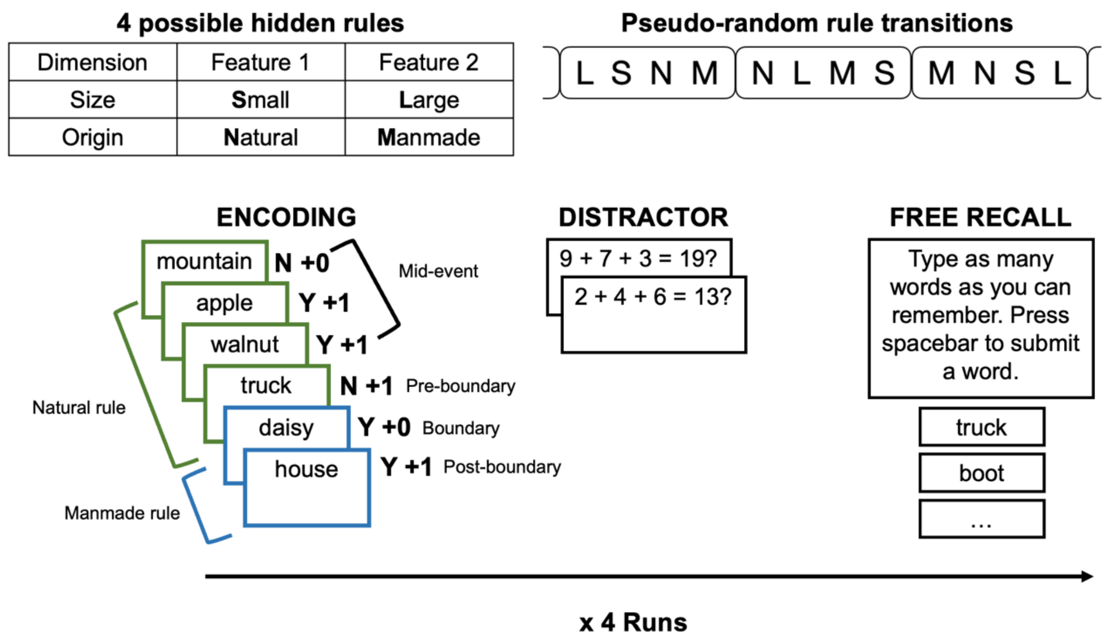


**Figure 1:** *Experimental design.* In the WRIT, participants indicate whether a word agrees with one of four hidden rules. The active rule repeats across events that last 6-8 trials. Then, it transitions pseudo-randomly such that each rule is experienced an equivalent number of times in each run. Here the word "daisy" marks the change from the natural rule (green) to the manmade rule (blue) (color added as aid, not shown to participants). 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.

# Methods and Materials

## Word rule inference task

We designed a novel word rule inference task (WRIT) inspired by a variant of a Wisconsin Card Sorting Task (Wilson and Niv 2012), 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 (Figure 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 (Unsworth and Miller, 2021). 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 minutes 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 (Howard and Kahana, 1999). 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). A further 8 were removed due to performance that was worse than chance on the distractor task. Finally, 1 participant was removed for not performing one of the recall runs. Our final sample was 73 participants (mean age = 30.27 years, 33 females). 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 (RL) provides a successful framework for understanding how people use reward learning to choose between stimuli (Sutton and Barto 2018). In this framework, participants are modeled as tracking the relative value of stimuli and using that to guide their decisions (Niv et al. 2015). In the current study, we developed an RL 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.

Specifically, on each trial, the model computes the expected value of both responses, as the sum of the expected rewards of each response under each rule weighted by $W$. It uses a standard softmax decision function to transform these values into a choice. The model then uses the reward feedback to update the weights of each rule. To do this, it first calculates a prediction error, which is the difference between the reward and the expected value of the given response. This prediction error is then weighted by a learning rate, and used to update the weights of each rule consistent with the choice. The weights of all rules that were inconsistent with the choice are slowly decayed towards zero (Niv et al. 2015; Rouhani and Niv, 2021).

We used *maximum a-posteriori* model fitting to estimate three free parameters for each participant separately. To account for differences in exploration between participants, we fit an inverse temperature parameter ($\beta$) for each. Lower values of this parameter indicate more exploratory behavior. We also fit a learning rate parameter ($\eta$) which determined the speed of updating of weights in response to new feedback. Finally, we fit a decay rate ($d$), which dictates how quickly the weights of rules not consistent with the choice decayed towards 0. Full details of the model and fitting procedure are in Supplemental Methods.

## 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 ($\mu = 79.2\%, \sigma = 12.7\%$). 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:

$$reward \sim position\_relative\_to\_boundary \times run + (1|participant)$$

For the run variable, the first run was treated as the reference, enabling us to examine the degree to which participants improved at the primary task across runs. For the trial position variable, the initial trial was treated as the reference.

### Response time

To examine how response times changed over the course of an event we ran a hierarchical mixed-effects model of the following form:

$$RT \sim trial\_within\_event \times 1(post\_boundary) + (1|\, participant) + (1|\, word)$$

Here, $\mathbf{1}(\boldsymbol{post\_boundary})$ indicates that post-boundary items are coded as 1, and all other items are treated as the reference case.

### Post-error slowing

As a measure of inference, we examined post-error response times relative to participant averages. We reasoned that trials following no reward would be slower than trials following reward. This would be indicative of a post-error slowing process (Danielmeier and Ullsperger 2011). To examine the presence of post-error slowing, we ran a hierarchical linear model of the following form:

$$RT \sim previous\_no\_reward \times boundary\_label + (1|\, participant) + (1|\, word)$$

Here, the boundary label variable is a categorical regressor, where position relative to the event boundary is coded according to the scheme introduced above**.**

## 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 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.

We measured temporal contiguity using a lag conditional response probability (CRP) approach (Kahana 1996). 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 CRP analyses were conducted using the psifr package (Morton, 2020).

### 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 (Michelmann et al. 2023). 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. 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 the nearest boundary. This was done for all changes between events and normalized within-subject. This provided us with relative proportions that each participant used boundary items as anchor points.

*Clustering analyses.* 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). The Supplemental Methods describes this analysis in more detail.

### Hierarchical mixed effects models predicting memory

After fitting the RL model, we used the best fitting parameters for each subject to extract reward prediction error (RPE) values for each trial they experienced. We then decomposed the RPE into its magnitude and sign and used these as separate regressors in a hierarchical mixed effects model:

$$recall\ success \sim RPE_{abs} \times RPE_{sign} + (1|participant) + (1|word)$$

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 is mean centered.

## 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" (Figure 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.

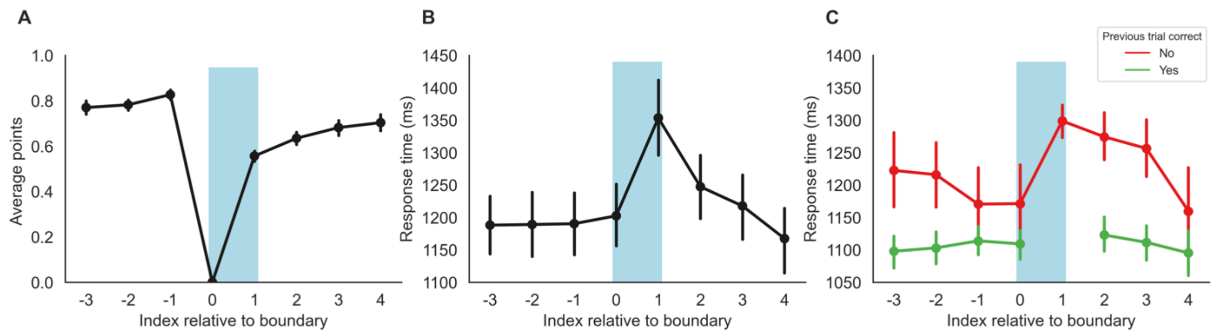


**Figure 2: Rule shifts reliably induce event boundaries. A.** Task performance was high before a rule shift, zero by definition at the boundary, and then increased over the next few trials as participants discovered the active rule. **B.** Participants were slowest at the post-boundary position, as 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 (green) 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. The shaded blue box indicates the boundary area.

**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 (Figure 2). A linear 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 ($\beta_{position\_relative\_to\_bound} = -0.11,\ CI95\% = [-0.15, -0.08],\ p < 0.001$) (Figure 2A). Participant performance also improved across runs ($\beta_{run} = 0.047,\ CI95\% = [0.02, 0.08],\ p < 0.001$). We also observed an interaction between run and position within the event ($\beta_{position\_relative\ X\ run} = 0.02,\ CI95\% = [0.01, 0.03],\ p < 0.001$). These results suggest that the error caused by the rule shift required participants to update their working event model.

Another hallmark of event segmentation is post-boundary slowing (Zwaan and Radvansky, 1998; Heusser et al. 2018). Consistent with this phenomenon, we found reliable increases in RT after rule shifts. A linear mixed effects model showed that, even though participants' responses became faster over the course of an event ($\beta_{trial\_within\_block} = -12.39,\ CI95\% = [-16.29, -8.48],\ p < 0.001$), they slowed down after the rule shift ($\beta_{post\_boundary} = 119.4,\ CI95\% = [96.33, 142.46],\ p < 0.001$) (Figure 2B). Because we defined the boundary as the first error in a new event, this result should be interpreted with the caveat that all trials following a rule shift involved an error on the previous trial. Thus, this effect may simply reflect post-error slowing. Therefore, we ran a linear mixed effects model to investigate post-error slowing for all positions in the event. This model found that errors induced slowing for all positions ($\beta_{prev\_no\_reward} = 76.70,\ CI95\% = [21.28, 132.13],\ p = 0.007$), but there was exaggerated slowing following the rule shift even while accounting for previous reward ($\beta_{label:post\_boundary} = 111.95,\ CI95\% = [57.66, 166.25],\ p < 0.001$) (Figure 2C).

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

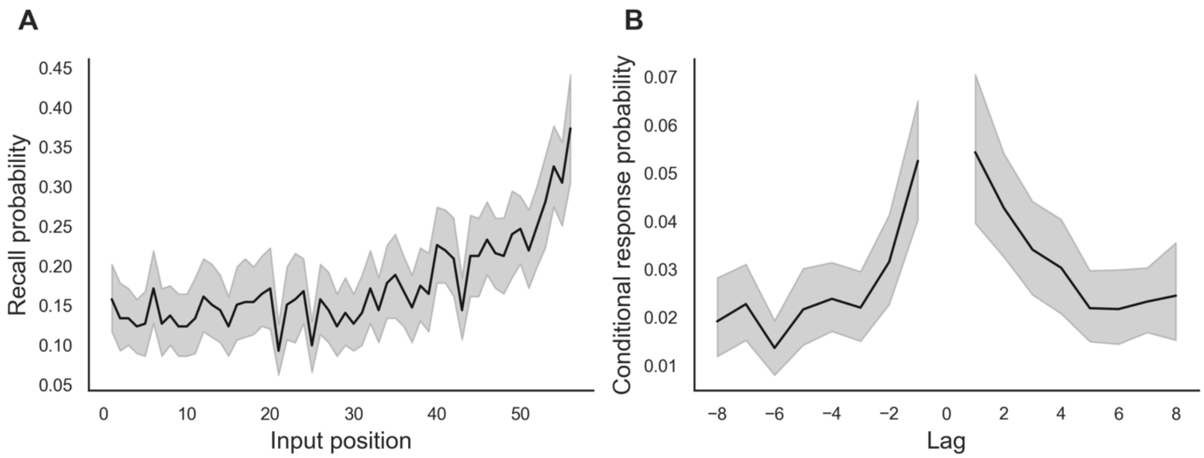


**Figure 3: General recall characteristics**. **A.** Serial position curve depicting the probability of a work being recalled as a function of its serial position. There was a recency effect where the final positions are better recalled than those in the middle, but no primacy effect. **B.** Conditional response probability curve showing evidence of canonical temporal contiguity effects in recall. Shading is 95% confidence interval.

### 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 16.6% of the words they encountered ($\mu = 0.166, \sigma = 0.072$). We observed strong evidence for a recency effect (Figure 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 (Healey 2018). Recall performance also featured a high degree of temporal contiguity (Figure 3B), with participants tending to recall sequences of words they experienced close in time in the WRIT (Kahana, Diamond, and Aka, 2022). Thus, overall, characteristics of the recall behavior observed in this study did not deviate from expected patterns.

## Impaired recall of post-boundary items

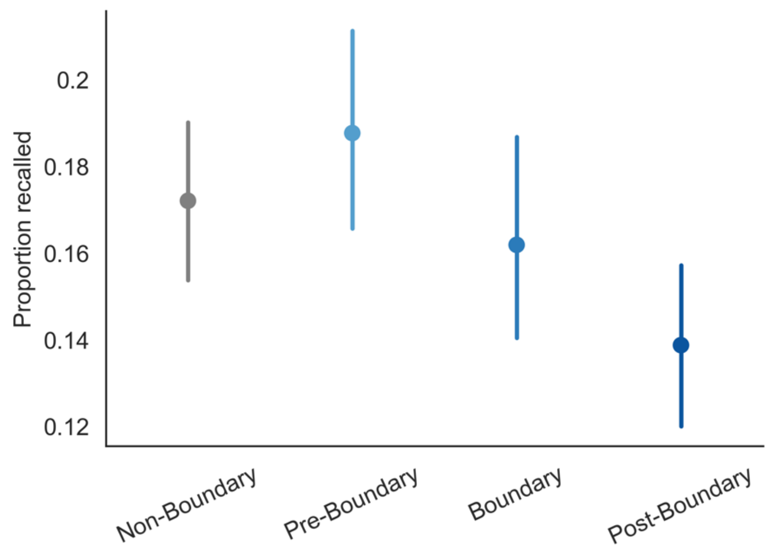


**Figure 4:** *Post-boundary inference leads to a recall deficit.* Words in the post-boundary position were remembered worse than all other positions. Interestingly, we found no recall benefit for boundary items. There is no difference between non-boundary, pre-boundary, or boundary items. Error bars represent 95% confidence intervals.

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 (Richmond and Zacks 2017). Boundary items were not recalled more often than non-boundary items ($t(72) = -0.80, Cohen's\ d = -0.15, p = 0.429$), nor were they recalled more often than pre-boundary items ($t(72) = -0.98$ , $Cohen's\ d = -0.20, p = 0.332$). Instead, contrary to the predictions of EST, we found evidence that event boundaries impaired memory. Specifically, post-boundary items were recalled less often than items in all other positions ( post-boundary against non-boundary $t(72) = -3.95, Cohen's\ d = -0.71, p < 0.001$, post-boundary against pre-boundary $t(72) = -3.21, Cohen's\ d = -0.643, p < 0.001$, and post-boundary against boundary $t(72) = -2.38, Cohen's\ d = -0.44, p = 0.020$; Figure 4).

## Boundaries and events structure recall

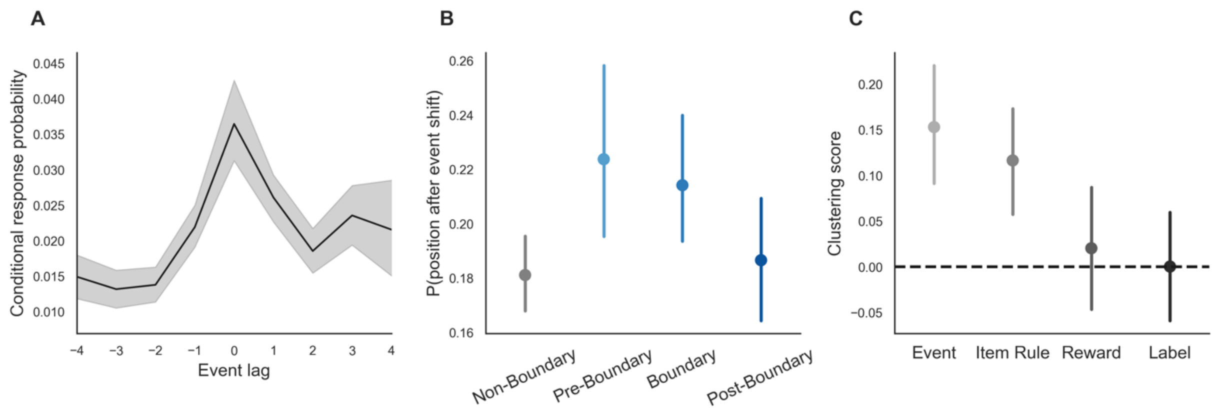


**Figure 5: Events organize free recall. A.** We calculated a CRP curve using the event lag to examine the degree of temporal contiguity while recalling events. Participants were most likely to recall words from the same event as the previously recalled word. **B.** The probability of recalling a word at a given position after transitioning events. Interestingly, pre-boundary items were most commonly used as the index into an event. **C.** Using the ARC 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.

### Events scaffold the structure of free recall

We next tested the extent to which event structure influenced free recall. In line with previous work (Michelmann, Hasson, and Norman 2023), we reasoned that participants would display temporal contiguity effects at the level of events. To this end, we calculated the CRP based not on serial position, but rather on which event an item belonged to (Figure 5A). Unlike a standard CRP 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 ($mean\ probability = 0.037$). Overall, forward across-event transitions were more probable than backward ones (test of negative lags vs positive ones $t(72) = 5.94, Cohen's\ d = 1.03, p < 0.001$. Meanwhile, we found no evidence for forward asymmetry for the next event, though this could be due to a lack of power to test this specific comparison (test of lag 1 vs lag -1 $t(72) = 1.52, Cohen's\ d = 0.26, p = 0.134$). In sum, these analyses revealed a tendency to stay within an event during recall, but 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 (Michelmann, Hasson, and Norman 2023), 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. In line with this prediction, items at boundaries were used as anchors more often than non-boundary items ($t(72) = 2.53, Cohen's\ d = 0.46, p = 0.014$) (Figure 5B). Somewhat unexpectedly, we found that pre-boundary items also anchored recall more often than non-boundary items ($t(72) = 2.06\ Cohen's\ d = 0.39, p = 0.043$). People did not differ in their tendency to use either pre-boundary or boundary items as anchors when transitioning between events ($t(72) = 0.14, Cohen's\ d = 0.02, p = 0.890$). Though pre-boundary and boundary items numerically anchored recall more so than post-boundary items, they did not differ significantly (pre-boundary: $t(72) = 1.41, Cohen's\ d = 0.30, p = 0.161$; boundary: $t(72) = 1.61, Cohen's\ d = 0.32, p = 0.112$). 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 (ARC) to examine the nature of recall clustering regardless of transition order. Using the ARC approach, we found that participants' recall was not only clustered by the event in which that word was experienced ($t(72) = 4.53, Cohen's\ d = 0.53, p < 0.001$), but also by the hidden rule that was active when the item was presented ($t(72) = 3.91, Cohen's\ d = 0.46, p < 0.001$) (Figure 5C). We found no evidence that participants used reward as a dimension to organize their free recall ($t(72) = 0.60, Cohen's\ d = 0.07\ p = 0.548$). 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 ($t(72) = 0.01, Cohen's\ d = 0.001, p = 0.993$).

## A reinforcement learning model predicts participants' recall

Given the importance of prediction errors in event boundary formation, we quantified the relative prediction error on each trial using an RL approach (Niv et al. 2015, Sutton and Barto 2018, Rouhani and Niv, 2021). According to EST, prediction errors in ongoing experience elicit event boundaries. Thus, we predicted that participant RPEs would align with our markers for event boundaries. As expected, the shift from one latent rule to another generated a large negative RPE (Figure 6A). The other trials consistently had low positive RPEs because the reward expectations incorporate the value of inactive rules that decay towards zero (see Methods: RL model).

Following previous work, we predicted that the absolute magnitude of RPEs would be positively correlated with increased memory (Rouhani and Niv, 2021; Rouhani et al. 2020). To assess this, we fit a hierarchical linear mixed-effects model predicting memory from both the magnitude and the sign of the RPE. Interestingly, we found that the magnitude of the RPE was not predictive of memory ($\beta_{RPE-abs} = -0.04,\ CI95\% = [-0.02, 0.09],\ p = 0.243$). The valence of the RPE was not a significant predictor for item recall, ($\beta_{RPE-val} = 0.03,\ CI95\% = [-0.2, 0.13],\ p = 0.663$). Importantly, we observed an interaction between magnitude and valence $\beta_{RPE-abs \times RPE-val} = -0.24,\ CI95\% = [-0.42, -0.06],\ p = 0.009$). Specifically, recall probability was positively related to the magnitude of negative RPEs and negatively related to the magnitude of positive RPEs (Figure 6B). Thus, the larger a participant's experienced negative RPE for an item, the more likely they were to remember it. In our task, more negative RPEs tend to occur when participants place a high value on a particular rule, and then receive negative feedback when responding according to that rule. According to our results, this form of corrective feedback makes them more likely to remember the item. Meanwhile, particularly positive RPEs occur when people are not placing a high enough value on a particular hypothesis. We found that this form of feedback leads to worse memory recall. Together these model-based findings are consistent with the more detailed memory analysis reported above: when participants have not found the active rule (large positive RPEs), subsequent item memory declines, but once they are more certain about the current rule (right before large negative RPEs), item memory improves.

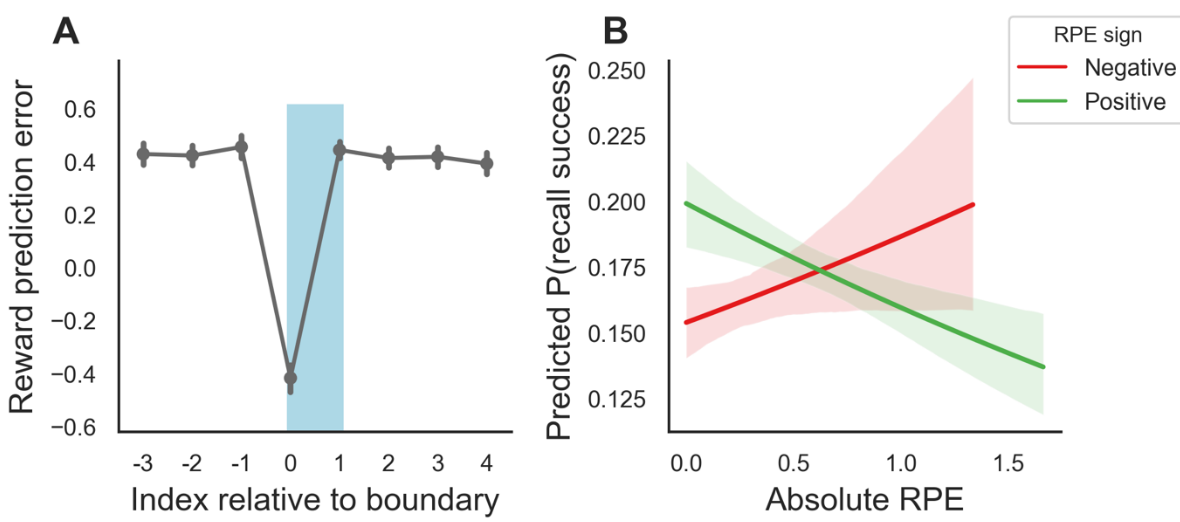


**Figure 6:** *Reinforcement learning predicts recall*. **A.** Across the task, participants generally experience small positive RPEs. This happens because people improved across the duration of an event. Moreover, each trial provides RPEs for both a relevant and irrelevant stimulus dimension. As the 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.** Predicted probability of recall success changes as a function of the magnitude and valence 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, show the opposite effect such that larger negative RPEs enhance the likelihood of recalling an item. (Error bars and shading are 95% confidence interval)

## 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 (Zacks et al. 2007). 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. On the other hand, 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 (Niv et al. 2015; Song et al. 2022). However, in contrast to previous studies (Rouhani et al. 2020; Rouhani and Niv 2021), we found an interaction between sign and magnitude of the RPE. Only negative RPEs and smaller positive RPEs were associated with greater recall success.

The influence of event boundaries on episodic memory has been the subject of much investigation (Swallow, Zacks, and Abrams 2009; Heusser et al. 2018; Morse et al. 2023). In recent work, Richmond and Zacks (2017) posit that event segmentation and event model construction processes result in increased encoding for items immediately surrounding an event boundary (see also Clewett et al., 2019). In this study, however, we saw no increase in memory for the pre-boundary or boundary items relative to the non-boundary items. This may result from a difference between interactively discovering an event boundary versus passively viewing it occur. 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 (Kool & Botvinick 2018). 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). 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 (Polyn, Norman, and Kahana 2009). Here, we found that higher-order event structure, delineated by shifting rules, served 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 previous work 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. Like many real-world experiences, this leads to ambiguity about when event boundaries take place. 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 (Rouhani et al. 2018, Rouhani and Niv, 2021). Here, we found a dissociation between positive and negative RPE effects on memory, namely that large positive RPEs dampen memory and large negative RPEs enhance memory. This may be due to the structure of our task, where the largest magnitude negative RPEs occur at the event boundary by design. Nonetheless, this reveals that memory is not unilaterally affected by surprise, a distinction not made by theories of event cognition. Further, we found that the direction of surprise is an important determinant of episodic memory formation. That is, undervaluing a prediction relative to the observed outcome may lead to poorer episodic encoding, whereas correction of an overvalued prediction may lead to enhanced episodic encoding.

A major 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, much 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. While our paradigm is less naturalistic than some prior studies of event memory that involve movie viewing (Swallow, Zacks, and Abrams 2009; Michelmann, Norman, and Kahana 2023), 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 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. One limitation of all free recall experiments, of course, is that participants tend to recall only a fraction of the studied list. This is a limitation in our current experimental paradigm as well. However, future studies can mitigate this limitation by using a recognition memory paradigm which would enable one to gauge memory signal for each item.

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 encoding of post-boundary items. This runs contrary to predictions of EST, as well as several empirical findings from studies using passive tasks. Finally, we found that prediction errors during encoding did not have a homogeneous influence on later recall. Specifically, we found reduced memory for items that were experienced with larger positive prediction errors, but increased memory when they were experienced with larger negative 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 novel theoretical and computational models of event cognition.

## 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.

## 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.

## References


Bailey, H. R., Kurby, C. A., Sargent, J. Q., & Zacks, J. M. (2017). Attentional focus affects how events are segmented and updated in narrative reading. *Memory & Cognition*, *45*(6), 940–955. https://doi.org/10.3758/s13421-017-0707-2

Barceló, F. (2021). A Predictive Processing Account of Card Sorting: Fast Proactive and Reactive Frontoparietal Cortical Dynamics during Inference and Learning of Perceptual Categories. *Journal of Cognitive Neuroscience*. https://doi.org/10.1162/jocn_a_01662

Danielmeier, C., & Ullsperger, M. (2011). Post-Error Adjustments. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00233

Healey, M. Karl. 2018. "Temporal Contiguity in Incidentally Encoded Memories." *Journal of Memory and Language* 102 (October): 28–40. https://doi.org/10.1016/j.jml.2018.04.003.

Heusser, Andrew C., Youssef Ezzyat, Ilana Shiff, and Lila Davachi. 2018. "Perceptual Boundaries Cause Mnemonic Trade-Offs Between Local Boundary Processing and Across-Trial Associative Binding." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 44 (7): 1075–90. https://doi.org/10.1037/xlm0000503.

Horwath, Elizabeth A., Nina Rouhani, Sarah DuBrow, and Vishnu P. Murty. 2023. "Value Restructures the Organization of Free Recall." *Cognition* 231 (February): 105315. https://doi.org/10.1016/j.cognition.2022.105315.

Howard, M. W., & Kahana, M. J. (2002). A Distributed Representation of Temporal Context. *Journal of Mathematical Psychology*, *46*(3), 269–299. https://doi.org/10.1006/jmps.2001.1388

Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition*, *24*(1), 103–109. https://doi.org/10.3758/BF03197276

Kahana, Michael J., Nicholas B. Diamond, and Ada Aka. 2022. "Laws of Human Memory." https://psyarxiv.com/aczu9/.

Michelmann, Sebastian, Uri Hasson, and Kenneth A. Norman. 2023. "Evidence That Event Boundaries Are Access Points for Memory Retrieval." *Psychological Science*, January, 09567976221128206. https://doi.org/10.1177/09567976221128206.

Morton, Neal W. 2020. "Psifr: Analysis and Visualization of Free Recall Data." *Journal of Open Source Software* 5 (54): 2669. https://doi.org/10.21105/joss.02669.

Morse, S. J., Karagoz, A., & Reagh, Z. (2023). *Event boundaries directionally influence item-level recognition memory* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/h8bj2

Niv, Yael, R. Daniel, A. Geana, S. J. Gershman, Y. C. Leong, A. Radulescu, and R. C. Wilson. 2015. "Reinforcement Learning in Multidimensional Environments Relies on

Attention Mechanisms." *Journal of Neuroscience* 35 (21): 8145–57. https://doi.org/10.1523/JNEUROSCI.2978-14.2015.

Pu, Y., Kong, X.-Z., Ranganath, C., & Melloni, L. (2022). Event boundaries shape temporal organization of memory by resetting temporal context. *Nature Communications*, *13*(1), 622. https://doi.org/10.1038/s41467-022-28216-9

Polyn, Sean M., James E. Kragel, Neal W. Morton, Joshua D. McCluey, and Zachary D. Cohen. 2012. "The Neural Dynamics of Task Context in Free Recall." *Neuropsychologia* 50 (4): 447–57. https://doi.org/10.1016/j.neuropsychologia.2011.08.025.

Polyn, Sean M., Kenneth A. Norman, and Michael J. Kahana. 2009. "A Context Maintenance and Retrieval Model of Organizational Processes in Free Recall." *Psychological Review* 116 (1): 129–56. https://doi.org/10.1037/a0014420.

Radvansky, Gabriel A., and Jeffrey M. Zacks. 2017. "Event Boundaries in Memory and Cognition." *Current Opinion in Behavioral Sciences* 17. https://doi.org/10.1016/j.cobeha.2017.08.006.

Radulescu, A., Shin, Y. S., & Niv, Y. (2021). Human Representation Learning. *Annual Review of Neuroscience*, *44*(1), 253–273. https://doi.org/10.1146/annurev-neuro-092920-120559

Richmond, Lauren L., and Jeffrey M. Zacks. 2017. "Constructing Experience: Event Models from Perception to Action." *Trends in Cognitive Sciences* 21 (12): 962–80. https://doi.org/10.1016/j.tics.2017.08.005.

Roenker, Daniel L., Charles P. Thompson, and Sam C. Brown. 1971. "Comparison of Measures for the Estimation of Clustering in Free Recall: Psychological Bulletin." *Psychological Bulletin* 76 (1): 45–48. https://doi.org/10.1037/h0031355.

Rouhani, Nina, and Yael Niv. 2021. "Signed and Unsigned Reward Prediction Errors Dynamically Enhance Learning and Memory." Edited by Thorsten Kahnt and Christian Büchel. *eLife* 10 (March): e61077. https://doi.org/10.7554/eLife.61077.

Rouhani, Nina, Kenneth A. Norman, and Yael Niv. 2018. "Dissociable Effects of Surprising Rewards on Learning and Memory." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 44 (9): 1430–43. https://doi.org/10.1037/xlm0000518.

Rouhani, Nina, Kenneth A. Norman, Yael Niv, and Aaron M. Bornstein. 2020. "Reward Prediction Errors Create Event Boundaries in Memory." *Cognition* 203: 104269. https://doi.org/10.1016/j.cognition.2020.104269.

Sargent, J. Q., Zacks, J. M., Hambrick, D. Z., Zacks, R. T., Kurby, C. A., Bailey, H. R., Eisenberg, M. L., & Beck, T. M. (2013). Event segmentation ability uniquely predicts event memory. *Cognition*, *129*(2), 241–255. https://doi.org/10.1016/j.cognition.2013.07.002

Shen, X., Ballard, I. C., Smith, D. V., & Murty, V. P. (2022). Decision uncertainty during hypothesis testing enhances memory accuracy for incidental information. *Learning & Memory*, *29*(4), 93–99. https://doi.org/10.1101/lm.053458.121

Sutton, Richard S., and Andrew G. Barto. 2018. *Reinforcement Learning: An Introduction*. Second edition. Adaptive Computation and Machine Learning Series. Cambridge, Massachusetts: The MIT Press.

Swallow, Khena M., Jeffrey M. Zacks, and Richard A. Abrams. 2009. “Event Boundaries in Perception Affect Memory Encoding and Updating.” *Journal of Experimental Psychology: General* 138 (2): 236–57. https://doi.org/10.1037/a0015631.

Unsworth, N., & Miller, A. L. (2021). Encoding dynamics in free recall: Examining attention allocation with pupillometry. *Memory & Cognition*, *49*(1), 90–111. https://doi.org/10.3758/s13421-020-01077-7

Wang, Yuxi Candice, and Tobias Egner. 2022. “Switching Task Sets Creates Event Boundaries in Memory.” *Cognition* 221 (April): 104992. https://doi.org/10.1016/j.cognition.2021.104992.

Wilson, Robert, and Yael Niv. 2012. “Inferring Relevance in a Changing World.” *Frontiers in Human Neuroscience* 5. https://www.frontiersin.org/articles/10.3389/fnhum.2011.00189.

Zacks, Jeffrey M. 2019. “Event Perception and Memory.”

Zacks, Jeffrey M., Christopher A. Kurby, Michelle L. Eisenberg, and Nayiri Haroutunian. 2011. “Prediction Error Associated with the Perceptual Segmentation of Naturalistic Events.” *Journal of Cognitive Neuroscience* 23 (12): 4057–66. https://doi.org/10.1162/jocn_a_00078.

Zacks, Jeffrey M., Nicole K. Speer, Khena M. Swallow, Todd S. Braver, and Jeremy R. Reynolds. 2007. “Event Perception: A Mind-Brain Perspective.” *Psychological Bulletin* 133 (2): 273–93. https://doi.org/10.1037/0033-2909.133.2.273.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*(2), 162–185. https://doi.org/10.1037/0033-2909.123.2.162