The role of prior-event retrieval in encoding changed event features

Mary M. Hermann[1] · Christopher N. Wahlheim[2] · Timothy R. Alexander[2] · Jeffrey M. Zacks[1]

Abstract

When people experience everyday activities, their comprehension can be shaped by expectations that derive from similar recent experiences, which can affect the encoding of a new experience into memory. When a new experience includes changes—such as a driving route being blocked by construction—this can lead to interference in subsequent memory. One potential mechanism of effective encoding of event changes is the retrieval of related features from previous events. Another such mechanism is the generation of a prediction error when a predicted feature is contradicted. In two experiments, we tested for effects of these two mechanisms on memory for changed features in movies of everyday activities. Participants viewed movies of an actor performing everyday activities across two fictitious days. Some event features changed across the days, and some features violated viewers' predictions. Retrieval of previous event features while viewing the second movie was associated with better subsequent memory, providing evidence for the retrieval mechanism. Contrary to our hypotheses, there was no support for the error mechanism: Prediction error was not associated with better memory when it was observed correlationally (Experiment 1) or directly manipulated (Experiment 2). These results support a key role for episodic retrieval in the encoding of new events. They also indicate boundary conditions on the role of prediction errors in driving new learning. Both findings have clear implications for theories of event memory.

Keywords Change detection · Event cognition · Prediction error · Proactive interference · Reminding

Everyday environments are dynamic, and so navigating the tasks of daily living requires recognizing changes and adapting appropriately. For example, suppose you began taking an art class and found an efficient driving route to class—but then on the second class found your preferred route blocked by construction. You would likely experience a prediction error upon encountering the closed road. Recognizing and adapting to this change could help you update your memory and choose a different route for the next class. Here, we examine potential mechanisms that govern the retrieval of past events in the service of recognizing everyday changes and updating memory to incorporate new information.

Updating episodic memory to incorporate changes

Episodic memory research has conceptualized changes in terms of *interference* (Underwood, 1957). Interference theories predict that episodic changes will impair memory when multiple features are associated with a common context (Anderson & Neeley, 1996). In our construction example, the two routes are associated with the context of driving to class. Because of this shared context, memory for the new route may be impaired by *proactive interference*, which occurs when earlier memories compete with retrieval of new information. Learning the new route could also impair memory for the previous route due to *retroactive interference*, which occurs when new learning impairs existing memories (McGeoch & McDonald, 1931).

Although the negative consequences of interference are well-established, they are not universal. For example, *retroactive facilitation* has been observed in paired-associate learning when the same cue appeared with different associated responses (e.g., Barnes & Underwood, 1959) and when the same cue appeared with different responses, but participants were told about the differences prior to encoding (Bruce &

✉ Jeffrey M. Zacks
jzacks@wustl.edu

[1] Department of Psychological and Brain Sciences, Washington University in St. Louis, 419B Somers Family Hall, St. Louis, MO 63130, USA

[2] Department of Psychology, University of North Carolina at Greensboro, Greensboro, NC 27412, USA

Springer

Weaver, 1973; Robbins & Bray, 1974). Similar findings were also observed in discourse comprehension (Ausubel et al., 1957). Collectively, these findings suggest that events including shared and distinctive features can sometimes facilitate memory, perhaps through integration. Consistent with this, interference can be reduced by integrating features from a common context into an event model (e.g., Radvansky, 2005).

The memory-for-change (MFC) framework proposes a mechanism by which changed features can impair or improve memory (Jacoby et al., 2015; Wahlheim & Jacoby, 2013). This proposal was inspired by the *recursive reminding hypothesis* (Hintzman, 2004, 2010, 2011), entailing that current event features can trigger a reminding of similar past events, leading to encoding of *configural representations* that include the original features, the changed features, and the discrepancy. When reminding leads to such memory updating, this should lead to proactive facilitation. But if reminding happens without updating, the increased strength of the initial event due to retrieval increases proactive interference.

Support for these predictions has been shown in paired associate learning experiments by Jacoby, Wahlheim, and colleagues. For example, Wahlheim and Jacoby (2013) instructed participants to study two lists of cue–response pairs, including responses that changed from the first to second presentation (e.g., *knee–bend* to *knee–bone*). During List 2, participants indicated when they detected changed responses, which was a measure of reminding. At test, participants were shown cues (e.g., *knee*) and were asked to recall the most recent responses (e.g., *bone*) and indicate when they remembered changes. Memory for recent responses was enhanced when changes were detected in List 2 and recollected at test, whereas proactive interference occurred when changes were detected but not later recollected. These results support the role of change recollection in memory updating and have since been replicated across various contexts (Jacoby et al., 2015; Wahlheim, 2014, 2015; Wahlheim et al., 2019).

The dynamics of memory retrieval in comprehension

Studies using word pairs are limited because they afford little insight into the dynamics of memory retrieval in comprehension. A key function of memory is to enable predictions about upcoming events, and *prediction errors* can be a powerful cue for memory updating (for a review, see Sinclair & Barense, 2019). For example, Kim et al. (2014) proposed a context-based mechanism by which similar contexts and cues trigger automatic predictions, and when an item that was predicted fails to appear, this error weakens the previous representation. This process allowed for "pruning" of inaccurate episodic memories through prediction error. Similarly, Sinclair and

Barense (2019) proposed that prediction error leads memories to become more malleable and susceptible to interference. Using *incomplete reminders* to induce prediction error, they found that interrupting videos during reactivation destabilized memories and increased intrusions from interfering videos. Further, Sinclair et al. (2020) found that such prediction errors signaled the hippocampus to switch from making internal predictions to an encoding mode that incorporated new information into memory (see also Bein et al., 2019).

Characterizing the role of ongoing memory retrieval and prediction in comprehension requires a theoretical framework that describes these dynamics, and experimental methods that are sensitive to them. To provide a theoretical framework, Event Memory Retrieval and Comparison Theory (EMRC; Wahlheim & Zacks, 2019) brings together the MFC framework with an account of the temporal dynamics of event comprehension (Zacks et al., 2007). EMRC proposes that as people observe everyday events, features cue recollection of similar past events and those representations influence predictions about upcoming events. When predictions of repeated event features are accurate, event models are maintained. But when predicted repetition is inconsistent with current event features, errors spike, and this drives detection of changed features and event model updating. The memory representation formed subsequent to such updating encodes both events and the retrieval that united them in working memory.

Wahlheim and Zacks (2019) tested this account using an *everyday changes* paradigm. In this paradigm, participants watched movies of an actor performing everyday activities on two fictive days in her life. There were two versions of each activity differing on one central feature (e.g., the actor woke up to a *clock* or *phone* alarm). For some activities, the central feature changed between movies. On a final cued recall test, participants recalled the central feature from each activity in the second movie (Day 2) and indicated whether the activity changed between days. When participants recollected *that* the activity had changed, they were more likely to recollect *what* the changed feature had become. Failure to recollect change was associated with proactive interference. According to EMRC, successful change detection and recollection enabled encoding of features from both movies, retrieval of features from Day 1 that elicited prediction errors on Day 2, and the temporal relationship between features. Further, Stawarczyk et al. (2020) used pattern-based fMRI with this paradigm to test the role of prior-event retrieval in memory updating. They found that both neural and self-report measures of reinstatement during the initial phase of an event predicted successful memory updating for changed events. These findings are consistent with studies showing that neural reinstatement of prior memories while encoding competing information is associated with reduced interference (e.g., Chanales et al., 2019; Koen & Rugg, 2016; Kuhl et al., 2010).

The present experiments

We extended work on event memory updating in the everyday changes paradigm by attempting to replicate the relationship between change recollection and recall, while testing for prior-event retrieval and prediction error mechanisms. We did this by modifying the event changes paradigm to include an overt prediction task (Zacks et al., 2011). Participants watched two movies, and during Day 2 movies, activities stopped after initial segments that repeated from Day 1. When the movies were stopped, participants made overt predictions about upcoming features. Experiment 1 manipulated whether activities repeated or changed, which resulted in prediction errors for some but not all activities. Experiment 2 manipulated prediction error using a response-contingent procedure that determined whether endings repeated or changed based on predictions. Previous findings led us to expect positive associations between change recollection and cued recall. We tested the hypothesis that prediction errors would be uniquely associated with change recollection and cued recall. To foreshadow, prior-event retrievals before encoding changes were associated with enhanced subsequent memory, but there was no evidence for a unique role of prediction error. In the General Discussion, we discuss the possibility that the overt prediction measure may not be an effective assay of the relevant prediction mechanisms in online event comprehension.

Experiment 1

Participants viewed a movie depicting a fictitious day in the life of the actor (Day 1). They then watched a movie depicting a second fictitious day (Day 2). Each Day 2 movie included *repeated* activities that were identical to clips shown on Day 1 and *changed* activities that began the same as Day 1 activities, but ended differently. Day 2 movies also included *novel* activities that did not overlap with Day 1. After watching the first segment of each Day 2 activity, participants predicted which of two endings would occur (see Fig. 1). Two days later, participants returned for a second session and completed a cued recall test in which they attempted to recall Day 2 activity features, indicated whether those activities changed from Day 1, and, for activities they classified as changed, attempted to recall Day 1 features. If prediction error facilitates memory updating for changed activities by enabling integrative encoding, then both recall accuracy and change recollection should be enhanced when Day 1 endings are predicted. Please note that "session" refers to the separate experimental sessions in which participants were asked to come into the lab. "Day" refers to each movie viewed by participants, depicting a series of activities unfolding over a fictitious day in the life of the actor. In Experiment 1, the Day 1 and Day 2 movies were both presented during the first of two experimental sessions. This was followed by a period of approximately 48 hours, and then the participant was brought back for the second experimental session, which consisted of the cued recall task.

Method

Participants

No previous studies provided a basis for estimating an effect size of prediction error effects on memory updating in the current paradigm. But earlier work using the everyday changes paradigm to investigate memory updating found robust
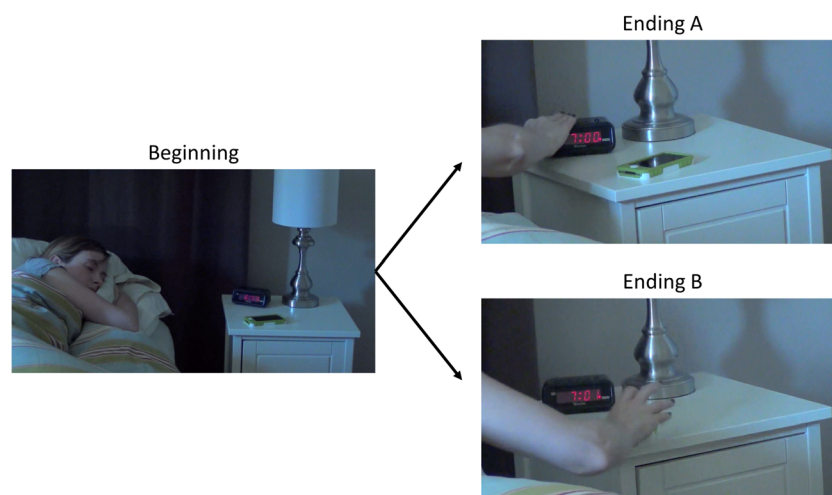

Fig. 1 Example of different endings for a changed activity. Images are from an activity in which the actor woke up to one of two alarms. The left image shows an image from the beginning segment that was the same for both versions of the activity, and the right images show the two criterial activity features corresponding with two different endings (i.e., an alarm clock in Ending A and a smart phone in Ending B)

results with 36 participants per experiment. Therefore, our stopping rule was to test at least 36 participants. We tested 43 students from the University of North Carolina at Greensboro across approximately one semester. Five dropped out, leaving 38 participants (21 females, M_{age} = 18.76, SD = 0.93, range: = 18–21 years). We conducted a sensitivity analysis based on our a priori hypothesis that prediction errors should lead to better recall and change recollection. According to G*Power Version 3.1.9.2 (Faul et al., 2009), 36 participants was enough to detect a medium-sized effect (d = 0.47) with power = .80, alpha = .05 (two-tailed). Participants received course credit.

Materials and design

The materials were two movies of a female actor performing daily activities around her home on two fictive days. There were two versions of each activity, differing in a central feature (see Fig. 1). Separate versions were created by dividing activity clips into two segments. The beginning segments were identical for each version (Duration: M = 10.07 s, range: 1–40 s) in duration. The ending segments included versions that differed on a central feature (Duration: M = 26.85 s, range: 2–88 s). To minimize item effects, all activity endings were congruent with larger activity schemas, and therefore should not be surprising given the beginnings.

We manipulated activity type within subjects to create repeated, novel, and changed conditions. Repeated activities appeared in both movies, novel activities appeared in the Day 2 movie only, and changed activities included one version in the Day 1 movie and the other version in the Day 2 movie. The material set included 59 total activities (45 critical, 14 filler activities). We counterbalanced the 45 critical activities by dividing them into three groups of 15 and rotating them through conditions. Activity changes were created by showing the same Day 2 movie to all participants and creating three formats of the Day 1 movie that varied on which activities had endings that differed from endings in the Day 2 movie.

Each of the Day 1 movies included 44 activities (30 critical, 14 fillers) and ranged in duration from 22 min and 58 s to 26 min and 33 s. The Day 2 movie included 59 activities (45 critical, 14 fillers) and had a total duration of 35 min 31 s. The Day 2 movie contained more critical items to allow for the comparison of novel activities. Activity type conditions in all movies appeared in a fixed-random order, with the constraint that no more than three critical activities from the same condition appeared consecutively. Filler activities repeated between movies and were included to make the movies more coherent and cohesive. This was necessary because the transitions between critical activities were sometimes discontinuous without intervening actions (e.g., the actor walking between the rooms in which she performed subsequent

activities). Memory for activity features was measured using a cued recall test that included cues for all 59 activities in the Day 2 movie. Test cues asked about the criterial feature of each activity (e.g., "What device awakened the actor?").

Procedure

All stimuli were presented using E-Prime 3 software (Psychology Software Tools, 2016) connected to a 24-in. monitor. The viewing distance was approximately 60 cm. Participants were tested individually in two sessions separated by 48 hours.

During the first session, participants watched both movies. Before viewing the Day 1 movie, participants were told to observe the actor and that they would be tested later; they also watched an example activity. Before viewing the Day 2 movie, participants were told how activity endings related between movies and to attend to those relationships. Participants were also told that the movie would stop intermittently and that they would be asked to predict activity endings; they were also shown a schematic of the prediction task. The Day 2 movie stopped after the first segment of each critical activity. Next, two images depicting central features from each possible activity ending appeared side-by-side. Participants predicted endings by pressing the "Q" or "P" key for the left or right image. Predictions were self-paced. A blank screen appeared for 5 s after each prediction during which participants mentally simulated the predicted ending. This was intended to generate robust prediction errors when actual and predicted activity endings differed. The Day 2 movie then resumed with the ending determined by the activity type condition. Participants were encouraged to note the relationship between the predicted and actual endings (see Appendix A for complete instructions).

We refer to predictions consistent with Day 1 endings as *memory-based*, and predictions inconsistent with Day 1 endings as *non-memory-based*. Novel activity predictions were a baseline measure of selecting each ending. The combination of activity and prediction types determined whether prediction errors occurred. Prediction errors occurred when changed endings followed memory-based predictions and repeated endings followed non-memory-based predictions. Correct predictions occurred when repeated endings followed memory-based predictions and changed endings followed non-memory-based predictions.

During the second session, participants completed the cued recall test. They were asked to recall Day 2 activity features, indicate which activities included changed endings, and recall Day 1 features when endings had changed. Before the cued recall test began, participants first watched an example changed activity comprising two clips with the same beginning and different endings, to illustrate what we considered to be changed activities. Then participants started the actual cued

recall test. Test cues asked about the central feature in associated activities (e.g., "What device awakened the actor?"), and participants typed their response. A prompt then asked whether the activity differed between the movies. Participants pressed the "1" or "2" key to indicate that activities changed or did not change. When participants indicated change, they were cued to recall the Day 1 feature.

Statistical approach

All analyses were conducted using R software (R Core Team, 2013). We took frequentist and Bayesian approaches. For frequentist analyses, we estimated probabilities for Day 2 recall and Day 1 intrusions from logistic mixed effects models including activity type and prediction type as fixed effects and subjects and items as random effects. We used models from the *glmer* function of the *lme4* package (Bates et al., 2015), and performed hypothesis tests using the *Anova* function of the *car* package (Fox & Weisberg, 2011), and post hoc comparisons using the *emmeans* function from the *emmeans* package (Lenth, 2019). The significance level was $\alpha = .05$.

For Bayesian analyses, we computed Bayes Factors (BF_{10}) using the *bayes_factor* function to compare models fitted with the *brm* function from the *brms* package (Bürkner, 2018) including random effects of subjects and items. Following earlier classifications (e.g., Jeffreys, 1961; Kass & Raftery, 1995), we interpreted support for the alternative hypothesis as weak ($BF_{10} = 1–3$), moderate ($BF_{10} = 3–10$), or strong ($BF_{10} > 10$), and support for the null hypothesis as weak ($BF_{10} = 1–.33$), moderate ($BF_{10} = .33–.10$), or strong ($BF_{10} < .10$). A BF_{10} of 1 indicates equal support for the null and alternative hypotheses.

Results

Cued recall response coding

Cued recall responses for Day 2 features were classified into four types. *Day 2 recall* included the central feature from the Day 2 movie. *Day 1 intrusions* included the central feature from the Day 1 movie (for repeated and novel activities, these were baseline estimates of reporting the central feature not from Day 2). *Ambiguous* responses included the action, but did not differentiate between the two possible central features. *Other errors* either did not refer to target actions or were omissions. We only had a priori hypotheses for Day 2 recall and Day 1 intrusions, so we did not analyze ambiguous responses and other errors.

Day 2 recall and Day 1 intrusions

We first examined overall Day 2 recall and Day 1 intrusions (Fig. 2, top panels) with separate models including the activity

type factor. For Day 2 recall (left panel), there was a significant effect, $\chi^2(2) = 22.46$, $p < .001$, with strong evidence for the alternative hypothesis, $BF_{10} = 9636.15$. Recall was significantly higher for repeated than novel, z ratio = 3.98, $p < .001$, and changed activities, z ratio = 4.28, $p < .001$, and was not significantly different between novel and changed activities, z ratio = 0.31, $p = .95$. For Day 1 intrusions (right panel), there was a significant effect, $\chi^2(2) = 29.19$, $p < .001$, with strong evidence for the alternative hypothesis, $BF_{10} = 2.19 \times 10^5$. Day 1 intrusions for changed activities were significantly greater than baseline intrusions for repeated, z ratio = 4.63, $p < .001$, and novel activities, z ratio = 4.42, $p < .001$, there was no significant difference for repeated and novel activities, z ratio = 0.21, $p = .98$.

Day 2 recall and Day 1 intrusions conditionalized on predictions

We then examined the association between predictions and memory accuracy using separate models for Day 2 recall and Day 1 intrusions including the activity type and prediction type factors (Fig. 3, top panels). We do not report effects of activity type redundant with those above. For Day 2 recall (left panel), there was no significant effect of prediction type, $\chi^2(1) = 2.13$, $p = .14$, with anecdotal evidence for the null hypothesis, $BF_{10} = 0.88$. There was no significant interaction, $\chi^2(2) = 1.31$, $p = .52$, with anecdotal evidence for the alternative hypothesis, $BF_{10} = 1.04$. To assay the association between prediction error and subsequent encoding uncontaminated by memory for Day 1 activities, we examined Day 2 recall for novel activities conditionalized on whether participants predicted the correct ending; there was no significant difference, $\chi^2(1) = 1.01$, $p = .32$, with anecdotal evidence for the null hypothesis, $BF_{10} = 0.99$. Finally, for Day 1 intrusions (right panel), there was no significant effect of prediction type, $\chi^2(1) = 0.39$, $p = .53$, with anecdotal evidence for the null hypothesis, $BF_{10} = 0.49$. There was no significant interaction, $\chi^2(2) = 1.13$, $p = .57$, with anecdotal evidence for the alternative hypothesis, $BF_{10} = 1.58$. Collectively, the results from the frequentist analyses indicated no significant effects, and the results from the Bayesian analyses provided no more than anecdotal evidence for the alternative or null hypotheses. These findings are therefore inconclusive regarding the direct association between prediction error and memory updating.

Day 2 recall benefits and memory-based predictions

Next, we tested the EMRC proposal that prior-event retrieval should enhance encoding of changed features by examining whether accurate retrieval of Day 1 features before changed Day 2 endings was associated with better Day 2 recall. We
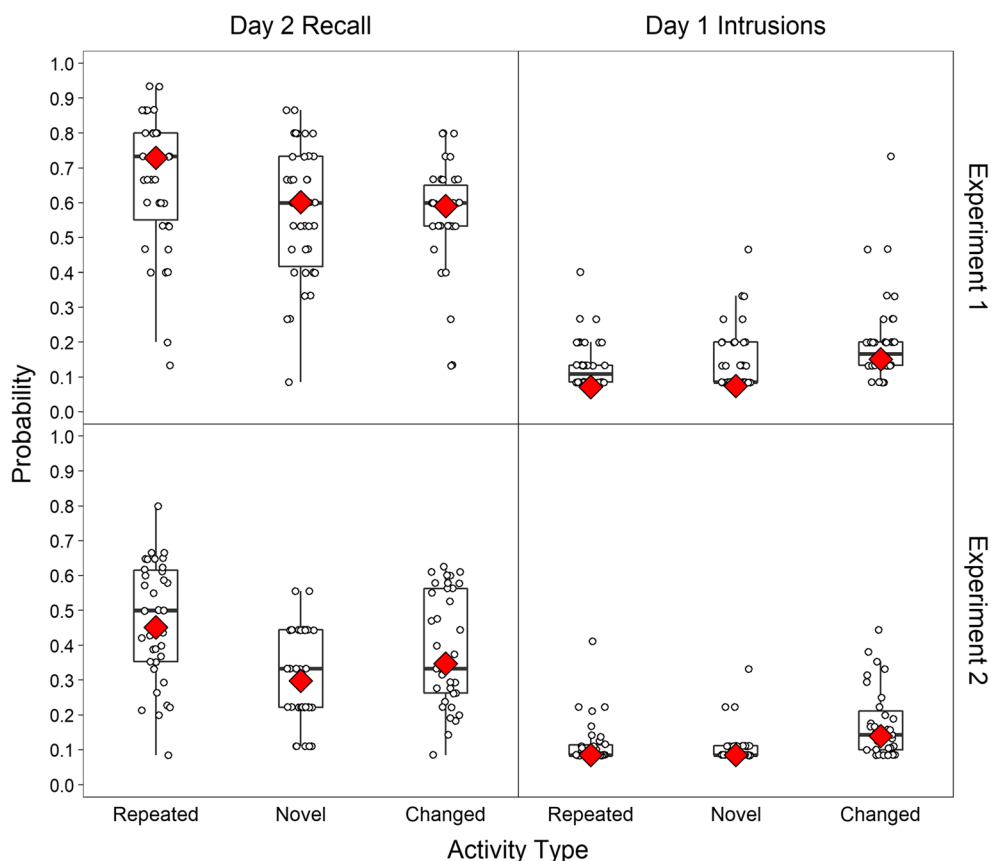
Fig. 2 Correct recall of Day 2 activities and intrusions of Day 1 activities: Experiments 1 and 2. In the box-and-whisker plots, the red diamonds indicate model-estimated probabilities, the horizontal lines indicate medians, the height of boxes mark the interquartile ranges, and the whiskers extend 1.5 times the interquartile ranges. Dots represent individual subject probabilities. (Color figure online)

assumed that retrieving Day 1 features would lead to predicted repetitions of those features, which we refer to as *memory-based predictions*. These contrast with *non-memory-based predictions*, that we assumed followed failures to retrieve Day 1 features. Since we did not instruct participants to predict Day 1 repetitions, we also did not predict a specific association between memory-based predictions and Day 2 recall. Consequently, we did not conduct hypothesis tests of that association. But when looking at the data post hoc, we found that participants made more memory-based predictions than would be expected by chance ($M = 66\%$, $SD = 17\%$), $t(37) = 5.80$, $p < .001$, and predicted endings for novel activities at chance ($M = 54\%$, $SD = 14\%$), $t(37) = 1.57$, $p = .06$. We therefore generated a post hoc hypothesis that participants varied in the extent to which they adopted a memory-based prediction strategy, and to the extent that participants relied relatively more on this strategy, memory-based predictions should be associated with better Day 2 recall.

We tested this hypothesis by computing for each participant the Day 2 recall benefit associated with memory-based predictions, operationalized as the difference in Day 2 recall between memory-based and non-memory-based prediction trials. We then correlated those difference scores with

participants' proportions of memory-based predictions (Fig. 4, left panel). Supporting our hypothesis, there was a significant positive association, $r(36) = .47$, 95% CI [.18, .69], $p < .01$, providing preliminary evidence for Day 1 retrieval benefits on recall of changed features. We also tested for differences in the correlations between memory-based predictions collapsed across all activities and Day 2 recall differences for repeated and changed activities for the 36 participants who had scores on each measure. We did this using the Williams's test of differences between dependent correlations using the *r.test* function in the *psych* package. There was no significant difference between correlations including recall differences for repeated, $r(34) = .33$, and changed, $r(34) = .08$, activities, $t = 1.06$, $p < .30$.

Change classifications conditionalized on prediction types

Table 1 (top row) displays the probabilities of classifying an activity as "changed" on the cued recall test. Responding "changed" is a correct response for changed activities, and an error for repeated and control activities. When changed activities were correctly classified, participants sometimes recalled the Day 1 features. Previous results indicate that

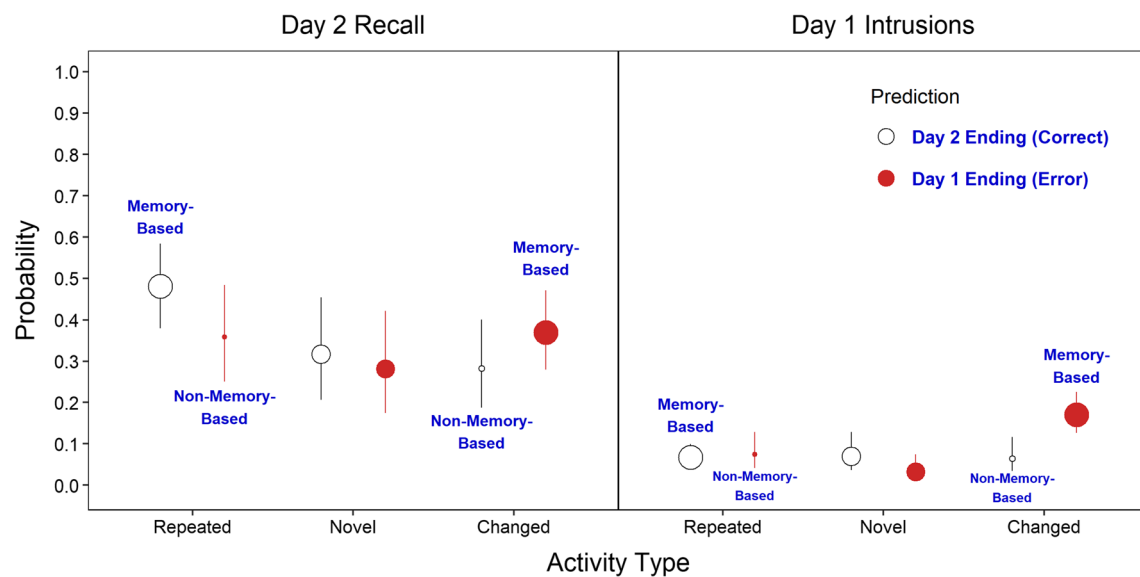## Experiment 1
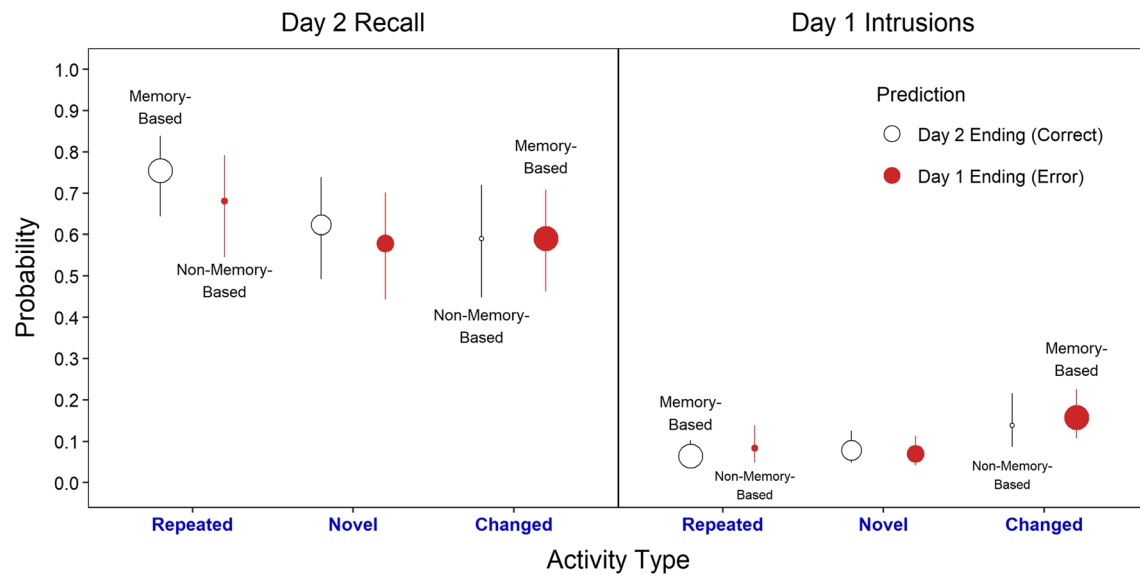
Experiment 2

Fig. 3 Correct recall of Day 2 activities and intrusions of Day 1 activities conditionalized on prediction error: Experiments 1 and 2. Boldfaced blue labels indicate conditions that were determined by the experimental design. Point areas indicate the relative proportions of observations in each cell. Error bars are 95% confidence intervals. (Color figure online)

classifying an activity as changed without being able to recall the Day 1 features is associated with poor memory for Day 2 features (Wahlheim & Zacks, 2019), so we further divided activities classified as "changed" based on recall of the central Day 1 feature (Table 2, top row). When changes were correctly classified, *change recollected* indicated when Day 1 features were recalled, and *change remembered but not recollected* indicated when Day 1 features were not recalled. Incorrect classifications of changed activities were categorized as *change not remembered*.

We tested the hypothesis that memory-based prediction errors should be associated with change recollection by conditionalizing classifications on prediction types (Table 2, second and third rows). The first column shows that changes were recollected significantly more often following predicted Day 1 endings, $\chi^2(1) = 6.49$, $p = .01$, $BF_{10} = 15.04$. The Bayes factor indicated strong evidence for the alternative hypothesis, suggesting that memory-based prediction errors were associated with better encoding of the fact that activity features had changed.
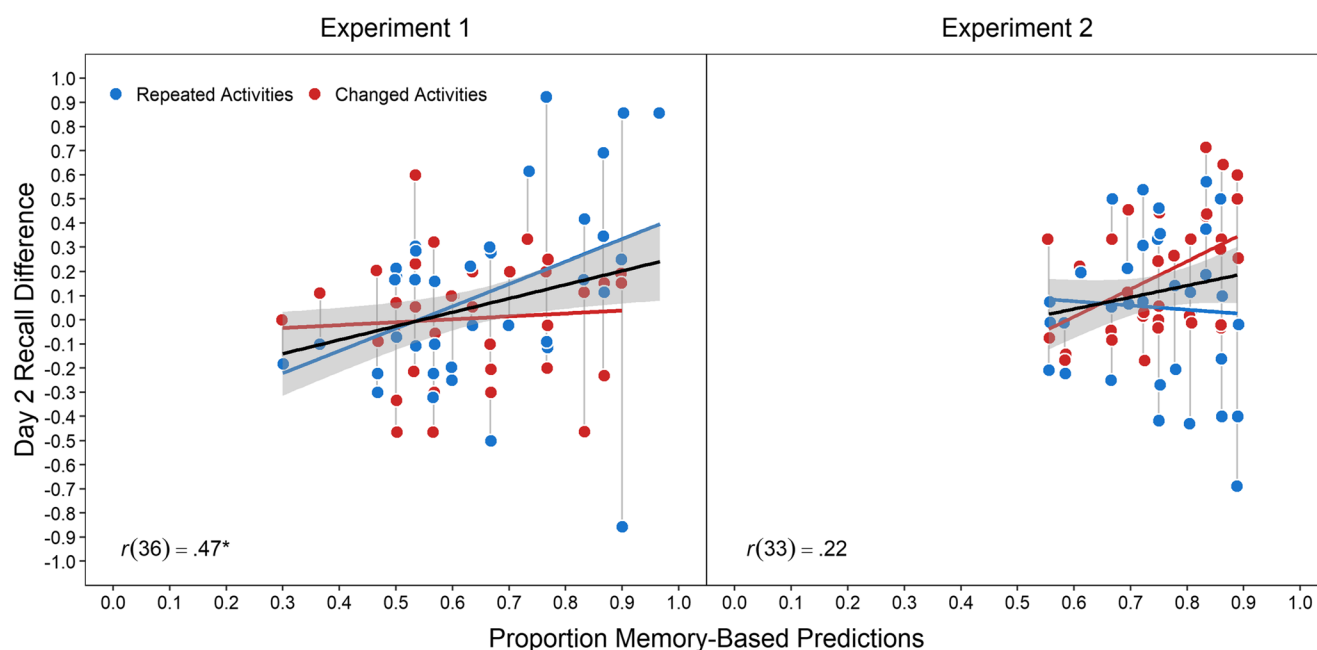
Fig. 4 Associations between memory-based predictions and corresponding recall benefits. Scatter plots showing the relationship between the proportion of memory-based predictions for each participant (*x*-axis) and the associated benefit on Day 2 recall of making a memory-based prediction (*y*-axis). Point pairs and connecting lines are individual participants. Single blue points without connecting lines in Experiment 1 (left panel) are participants who only made both memory-based and non-memory-based predictions for repeated activities. Best-fitting regression lines appear in black for correlations collapsed across activity types, in blue for correlations within repeated activities, and in red for changed activities. The shaded regions are 95% confidence intervals for correlations collapsed across activity types. *p < .01. (Color figure online)

Day 2 recall and Day 1 intrusions conditionalized on classifications and predictions

The results so far show equivocal support for the role of prediction error in memory updating for changed activities. There was inconclusive evidence for the main prediction of EMRC: that prediction error should be directly associated with enhanced Day 2 recall. But prediction error was clearly associated with enhanced change recollection, which, according to EMRC, should also be associated with enhanced Day 2 recall. To further understand this combination of associations, we conditionalized Day 2 recall and Day 1 intrusions for changed activities on both prediction types and change classifications. We expected to replicate earlier findings showing an

association between change recollection and enhanced Day 2 recall (e.g., Wahlheim & Zacks, 2019), but it was unclear whether that association would be comparable for both prediction types and how predictions would associate with Day 1 intrusions.

We fitted separate models to each recall measure including the change classification and prediction type factors (Fig. 5, top panels). The model for Day 2 recall (left panel) indicated a significant effect of change classification, $\chi^2(2) = 114.23$, $p < .001$, with strong evidence for the alternative hypothesis, $BF_{10} = 5.31 \times 10^{43}$; no significant effect of prediction type, $\chi^2(1) = 1.37$, $p = .24$, $BF_{10} = 1.34$, with only anecdotal evidence for the alternative hypothesis. The interaction was not significant, $\chi^2(2) = 1.86$, $p = .39$, but there was moderate evidence for the alternative hypothesis, $BF_{10} = 8.05$. Given that the only unambiguous effect was for change classification, we followed it up with pairwise comparisons. Day 2 recall was significantly higher when change was recollected than when change was remembered but not recollected and when change was not remembered, smallest *z* ratio = 7.75, $p < .001$. There was no significant difference between the latter classifications, for which change was not recollected, *z* ratio = 0.58, $p = .83$.

The model for Day 1 intrusions (right panel) did not include change recollection because those responses were rare and ambiguous to interpret. Although there was no significant effect of change classification, $\chi^2(2) = 2.01$, $p = .16$, there was moderate evidence for the alternative hypothesis, $BF_{10} = 4.80$. There was

Table 1 Model-estimated probabilities of change classifications as a function of activity type: Experiments 1 and 2

Experiment	Activity Type		
	Repeated	Novel	Changed
Experiment 1	.09 [.06, .13]	.29 [.23, .37]	.66 [.58, .73]
Experiment 2	.20 [.15, .26]	.27 [.20, .35]	.44 [.36, .52]

Note: Probabilities for changed activities are correct classifications, whereas probabilities for novel and repeated activities are incorrect classifications. 95% confidence intervals are displayed in brackets.

Table 2 Probabilities of change classifications as a function of classification and prediction type: Experiments 1 and 2

Experiment	Prediction Type	Classification Type		
		Change Recollected	Change Remembered Not Recollected	Change Not Remembered
Experiment 1	All activities	.51 [.46, .55]	.13 [.11, .16]	.36 [.32, .40]
	Day 2 ending (correct)	.43 [.36, .49]	.11 [.07, .15]	.46 [.39, .53]
	Day 1 ending (error)	.54 [.49, .60]	.16 [.13, .19]	.31 [.26, .36]
Experiment 2	All activities	.29 [.25, .32]	.14 [.10, .19]	.55 [.51, .59]
	Day 2 ending (correct)	.10 [.06, .15]	.15 [.10, .21]	.75 [.68, .81]
	Day 1 ending (error)	.35 [.31, .39]	.17 [.13, .20]	.48 [.44, .53]

Note: For prediction type, "all activities" included all observations in the changed conditions, "Day 2 ending (correct)" only included the observations when participants predicted the Day 2 ending (correct predictions), and "Day 1 ending (error)" only included the observations when participants predicted the Day 1 ending (prediction errors). 95% confidence intervals are displayed in brackets.

a significant effect of prediction type, $\chi^2(1) = 4.67$, $p = .03$, with strong evidence for the alternative hypothesis, $BF_{10} = 12.21$. These effects were qualified by a significant interaction, $\chi^2(1) = 8.56$, $p < .01$, for which there was strong evidence for the alternative hypothesis, $BF_{10} = 179.01$, indicating comparable intrusions when change was remembered, z ratio $= 1.33$, $p = .55$, and more intrusions following prediction errors when changes were not remembered, z ratio $= 3.36$, $p < .01$.

Collectively, the results from these conditional analyses partly support EMRC. Change recollection was associated with enhanced Day 2 recall. This benefit was comparable for both prediction types, and change was recollected more following predicted Day 1 endings. But when changes were not remembered, prediction errors were associated with more memory errors (i.e., Day 1 intrusions). These results suggest that prediction errors may have enhanced Day 2 recall by
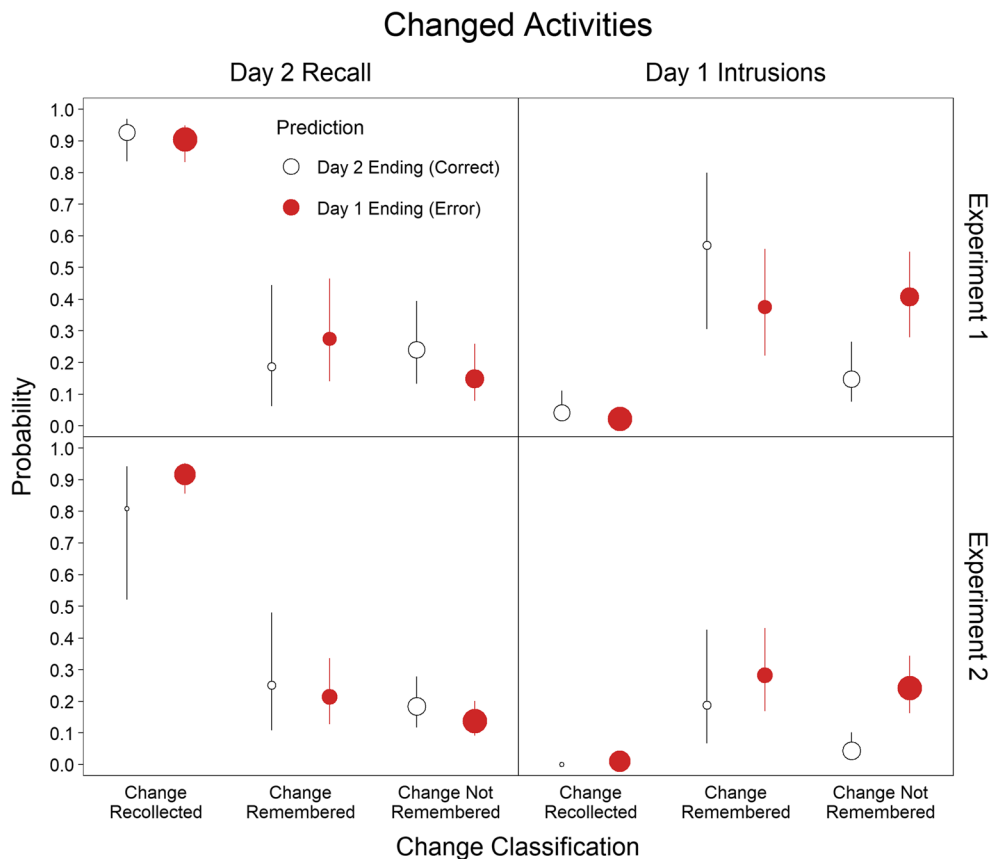

Fig. 5 Correct recall of Day 2 activities and intrusions of Day 1 activities conditionalized on prediction error and change classifications: Experiments 1 and 2. All correct predictions were non-memory-based, and all prediction errors were memory-based. Point areas indicate the relative proportions of observations in each cell. Error bars are 95% confidence intervals

enabling integrative encoding, but such enhancement was off-set in summary scores by the increase in memory errors observed when recall subsequent to memory-based predictions was not recollection-based.

Discussion

Changing a central activity feature was associated with enhanced Day 2 recall when the change and Day 1 feature were both remembered (i.e., when change was recollected). But when changes were not recollected, memory accuracy was impaired. These findings are consistent with EMRC predictions and earlier findings (Wahlheim & Zacks, 2019). The more novel contribution of this experiment was to test the EMRC proposal that successful retrieval of past events leads to prediction errors that improve encoding of changes. Although we did not instruct participants to use their Day 1 memories to predict Day 2 endings, exploratory analyses suggested that many participants did, and for those participants, making a memory-based prediction was associated with better subsequent recall of Day 2 features. This supports the idea that retrieval during learning, and even during active encoding, is associated with enhanced memory updating. In Experiment 2, we further examined this retrieval effect more directly by instructing participants to use Day 1 memories to form predictions during Day 2 viewing.

We hypothesized that memory-based prediction errors should be associated with enhanced Day 2 recall; estimates of correct recall and intrusions failed to support or to strongly reject this hypothesis. The association between pure prediction error and Day 2 recall of novel activities was also inconclusive. The absence of a direct association between prediction errors and recall for changed activities can be explained as reflecting a balance of offsetting enhancement and impairment that depended on recollection-based retrieval. Memory-based predictions were associated with more frequent change recollection, which was associated with enhanced Day 2 recall. But these predictions were also associated with more Day 1 intrusions when changes were not remembered (i.e., no recollection-based retrieval). In Experiment 2, we further examined the consequences of this balance on recall of changed activities by directly manipulating the experience of prediction error.

Experiment 2

In Experiment 2, we made the experience of prediction error independent of one's ability to retrieve Day 1 features by making the ending shown after each prediction made during Day 2 viewing dependent on that prediction. Instead of manipulating activity types by preassigning endings to Day 2 activities, we manipulated prediction errors by using a response-contingent display. If an activity was assigned to the *no prediction error* condition, it was always followed by predicted ending; if an activity was assigned to the *prediction error* condition, it was always followed by the opposite ending than predicted. In other words, whereas in Experiment 1 each activity was randomly assigned to be repeated or changed (or novel) and prediction error depended on participants' predictions, in Experiment 2, prediction error or no prediction error was directly manipulated, and assignment of activities to repeat or change depended on viewers' responses. Also, to encourage participants to adopt a memory-based prediction strategy during Day 2 viewing, we instructed participants to make predictions based on memory for Day 1 endings.

We expected to replicate the overall patterns from Experiment 1. But we also expected that instructing participants to make memory-based predictions and the consequent response-contingent prediction errors would lead to more opportunities for enhanced memory updating associated with change recollection.

Method

Participants

Experiment 1 did not show an association between prediction errors and memory accuracy in summary scores. Thus, there are still no findings on which to estimate the effect size for prediction error benefits, so we followed the stopping rule from Experiment 1. We tested 42 participants from Washington University in St. Louis, but five dropped out, and data from two were lost due to technical errors. The final sample included 35 participants (23 females; M_{age} = 19.80 years, SD = 1.41, range: 17–24 years). Participants received either $10 per hour or course credit.

Materials and design

The materials and design were similar to Experiment 1. We used the same movies, but manipulated prediction type within subjects to create prediction errors. The Day 1 movies included 36 of the 45 critical activities that were either repeated or changed in the Day 2 movie (i.e., *recurring* activities) and 14 fillers that would repeat in the Day 2 movie (50 total activities). The Day 2 movie included all 59 activities. Of the 45 critical activities, 36 appeared in both movies as recurring activities, and nine appeared as novel activities only in the Day 2 movie.

We counterbalanced both the assignment of activities to prediction type conditions and the two versions of activities to the Day 1 movie by first dividing the two versions of the 36 recurring activities (72 total) into four groups of 18 activities. Then, we assigned each group to each condition once, thus

creating four experimental formats. The nine activities shown as novel clips on Day 2 also differed between these two formats and were divided similarly according to the manipulation. The nine novel activities also differed between the two formats so that activity was not confounded with activity type. The Day 1 movie format duration ranged from 21 min and 19 s to 22 min and 28 s. The total duration of the Day 2 movie depended on predictions, but could have ranged from 23 min and 57 s, if all predictions resulted in the shorter ending, to 26 min and 50 s, if all predictions resulted in the longer ending. There were differences in the clip lengths, with the average difference between clip versions being 1.13 seconds.

Procedure

All experimental stimuli were presented using PsychoPy software (Peirce et al., 2019) on an Apple iMac computer with a 21.5-in. screen. The viewing distance was approximately 60 cm. Participants were tested individually in three sessions, each separated by exactly 7 days. The additional delay between movies relative to Experiment 1 was to prevent participants from always being able to make memory-based predictions (see Appendix B for task instructions).

During the first session, participants watched the Day 1 movie and were told to attend to the actor's actions and that they would be tested later. Participants watched two practice activities and then watched the Day 1 movie as one continuous film.

During the second session, participants watched the Day 2 movie and were told how the upcoming recurring activities would relate to those in the Day 1 movie, but the instructions did not include a description of novel activities. Next, participants practiced the prediction task twice. One example was a prediction error trial and the other was not. The experimenter explicitly indicated the repeated feature from Day 1 to Day 2 in the trial without prediction error and the changed features from Day 1 to Day 2 in the prediction error trial to highlight the type of change that participants were supposed to notice. Participants then performed the prediction task for each activity. The task was the same as in Experiment 1, except that participants were asked to predict what they thought would happen on Day 2 based on their memory for Day 1. This is a critical difference from Experiment 1, in which participants were not instructed about strategy use for the prediction task. Participants made their predictions by pressing either the "1" or "2" key to choose the left or right image, respectively.

The Day 2 ending for each activity was contingent upon participant predictions and the prediction type condition (for a schematic, see Fig. 6). On trials without prediction errors, participants viewed the activity version that matched their predictions. On trials with prediction errors, participants viewed the activity version opposite of their predictions (e.g., a *clock* alarm prediction followed by a *phone* ending). As in Experiment 1, predictions were described as either memory-based (consistent with Day 1 endings) or non-memory-based (inconsistent with Day 1 endings). This did not apply to novel activities. This task feature resulted in recurring activities being repeated or changed from Day 1.

During the third session, participants completed the cued recall test of activity features from both movies that also included change classification judgments. In contrast to Experiment 1, the test cues were the first segments of the Day 2 activities. We modified test cues to preclude ambiguity about which activities they signaled. After the cue appeared, participants were asked to type the ending of corresponding Day 2 activity. Participants were then asked to report whether the activity ending changed or repeated from Day 1 to Day 2. Participants pressed the "1" or "2" key to indicate that activities changed or repeated, respectively. When participants indicated that an activity changed, they were then prompted to type the Day 1 ending. Participants were also asked to classify novel activities as changed or repeated but we did not analyze these responses given that the classifications did not map on to the activity type.
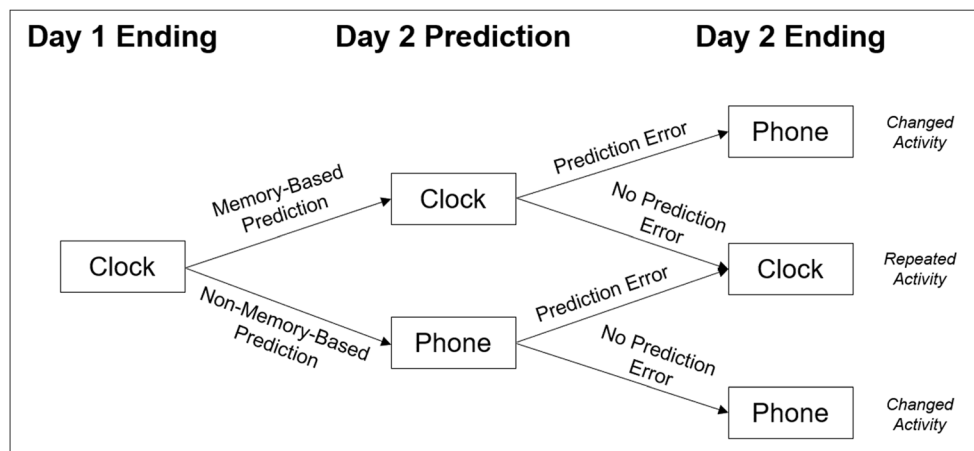

Fig. 6 Schematic illustrating prediction types in Experiment 2 using the example activity from Fig. 1

Results

Day 2 recall and Day 1 intrusions

As in Experiment 1, we first examined overall Day 2 recall and Day 1 intrusions (Fig. 2, bottom panels) with separate models including the activity type factor. For Day 2 recall (left panel), there was a significant effect, $\chi^2(2) = 16.16$, $p < .001$, with strong evidence for the alternative hypothesis, $BF_{10} = 474.81$. Recall was significantly higher for repeated than novel, z ratio = 3.38, $p = .002$, and changed activities, z ratio = 3.30, $p = .003$, and was not significantly different between novel and changed activities, z ratio = 1.18, $p = .47$. For Day 1 intrusions, there was a significant effect, $\chi^2(2) = 27.54$, $p < .001$, with strong evidence for the alternative hypothesis, $BF_{10} = 3.20 \times 10^5$. Day 1 intrusions for changed activities were significantly greater than baseline intrusions for repeated, z ratio = 4.17, $p < .001$, and novel activities, z ratio = 4.07, $p < .001$; repeated and novel activities were not significantly different, z ratio = 1.19, $p = .46$.

Day 2 recall and Day 1 intrusions conditionalized on predictions

We then examined the association between predictions and memory accuracy using separate models for Day 2 recall and Day 1 intrusions including activity type and prediction type as factors (Fig. 3, bottom panels). We do not report effects of activity type redundant with those above. For Day 2 recall (left panel), there was no significant effect of prediction type, $\chi^2(1) = 0.43$, $p = .51$, with anecdotal evidence for the null hypothesis, $BF_{10} = 0.40$. There was a significant interaction, $\chi^2(2) = 8.07$, $p = .02$, with strong evidence for the alternative hypothesis, $BF_{10} = 51.27$. Despite statistical support for the interaction, pairwise comparisons indicated no significant differences between prediction types within each activity type conditions, largest z ratio = 2.25, $p = .21$. Thus, the interaction reflected opposite numerical trends showing greater recall following predicted Day 2 endings for repeated and novel activities and greater recall following predicted Day 1 endings for changed activities. Tests of the association between "pure" prediction error and novel activity recall indicated no significant effect, $\chi^2(1) = 2.16$, $p = .14$, but moderate evidence for the alternative hypothesis, $BF_{10} = 3.93$.

The pattern of Day 2 recall for repeated and changed activities is consistent with the hypothesis that memory-based predictions should be associated with enhanced memory for Day 2 activities. We were able to directly test this hypothesis because we instructed participants to make their predictions based on memory for Day 1. As in Experiment 1, participants made more memory-based predictions than would be expected by chance ($M = 75\%$, $SD = 10\%$), $t(34) = 14.11$, $p < .001$, and predicted novel activity endings at chance ($M = 51\%$, SD

= 6%), $t(34) = 0.50$, $p = .31$. Analyses of memory-based predictions, which were applicable only to the repeated and changed activities, indicated a significant effect of activity type, $\chi^2(1) = 9.75$, $p < .01$, with strong evidence for the alternative hypothesis, $BF_{10} = 49.52$, a significant effect of prediction type, $\chi^2(1) = 7.96$, $p = .01$, with strong evidence for the alternative hypothesis, $BF_{10} = 20.00$, and no significant interaction, $\chi^2(1) = 0.09$, $p = .76$, with anecdotal evidence for the null hypothesis, $BF_{10} = 0.87$. These results show that memory-based predictions were directly associated with higher Day 2 recall regardless of prediction accuracy.

For Day 1 intrusions (right panel), there was a significant effect of activity type, $\chi^2(2) = 13.51$, $p < .01$, with strong evidence for the alternative hypothesis, $BF_{10} = 3.08 \times 10^5$, and no significant effect of prediction type, $\chi^2(1) = 2.61$, $p = .11$, with anecdotal evidence for the alternative hypothesis, $BF_{10} = 2.89$. A significant interaction, $\chi^2(2) = 9.37$, $p = .01$, with strong evidence for the alternative hypothesis, $BF_{10} = 257.89$, showed that for changed activities, there were significantly more intrusions associated with predicted Day 1 than Day 2 endings, z ratio = 3.28, $p = .01$, but the intrusion rates within repeated and novel activities were not significantly different, largest z ratio = 1.46, $p = .69$.

Although these results collectively show the same qualitative pattern as in Experiment 1, they also show that memory-based predictions were associated with significantly greater Day 2 recall and Day 1 intrusions. These results suggest that encouraging participants to predict Day 1 endings both improved and impaired memory updating. Below, we describe how these effects depended on participants' use of recollection-based retrieval.

Day 2 recall benefits and memory-based predictions

As described in Experiment 1, memory-based predictions occurred when participants predicted Day 1 endings. But here we instructed participants to predict Day 2 endings based on their memory for Day 1 endings. This appeared to increase memory-based predictions, as such predictions occurred more often in Experiment 2 than Experiment 1. The absence of instructions in Experiment 1 allowed maximal variability in participants' prediction strategy. This resulted in a positive association between memory-based prediction frequency and its benefits for Day 2 recall (see Experiment 1 for description of measures). Mandating memory-based predictions through instructions should reduce that correlation. Consistent with this hypothesis, Fig. 4 shows a nonsignificant association between these measures, $r(33) = .22$, 95% CI [−.12, .52], $p = .19$. We also tested for differences between correlations including recall differences for repeated and changed activities using the Williams' test. The correlation including recall differences for repeated activities, $r(33) = −.05$, was significantly different from the correlation including recall differences for changed

activities, $r(33) = .49$, $t = 2.68$, $p < .01$. We had no a priori reason to expect this difference and have no explanation for it.

Change classifications conditionalized on prediction type

As in Experiment 1, we estimated change classification probabilities using separate models for each type. The overall change classification probabilities (Table 1, bottom rows) are lower than in Experiment 1, likely because Experiment 2 included longer delays between each phase. Experiment 2 replicated the downstream consequences of predictions made during Day 2 on the various change classifications shown in Experiment 1 (Table 2, bottom two rows). Of particular interest, changes were recollected significantly more often following erroneous predictions of Day 1 endings, $\chi^2(1) = 22.69$, $p < .001$ (there was strong evidence for the alternative hypothesis, $BF_{10} = 4.05 \times 10^5$, which supported the hypothesis that memory-based prediction errors should improve encoding and recollection of changes).

Day 2 recall and Day 1 intrusions conditionalized on classifications and predictions

As in Experiment 1, we tested the predictions that memory-based prediction errors should improve encoding and recollection of activities and their order and lead to impaired memory when recollection fails. We did this by fitting models with change classification and prediction type as fixed effects to Day 2 recall and Day 1 intrusions (see Fig. 5, bottom panels).

Day 2 recall (left panel) showed a significant effect of change classification, $\chi^2(2) = 150.68$, $p < .001$, with strong evidence for the alternative hypothesis, $BF_{10} = 6.08 \times 10^{52}$, and no significant effect of prediction type, $\chi^2(1) = 0.35$, $p = .56$, with anecdotal evidence for the null hypothesis, $BF_{10} = 0.88$. Although the interaction was not significant, $\chi^2(2) = 2.65$, $p = .27$, there was strong evidence for the alternative hypothesis, $BF_{10} = 10.89$. We only followed the change classification effect with pairwise comparisons because it was the only unambiguous effect. Day 2 recall was significantly higher when change was recollected than when change was remembered but not recollected and when change was not remembered, *smallest z* ratio = 6.46, $p < .001$. There was no significant difference when change was remembered but not recollected and when it was not remembered, *z* ratio = 1.40, $p = .34$. These results replicate the strong positive association between change recollection and Day 2 recall.

For Day 1 intrusions (right panel), we again excluded change recollection observations from the models due to their rarity and ambiguity. There was no significant effect of change classification, $\chi^2(1) = 2.01$, $p = .16$, with moderate support for the alternative hypothesis, $BF_{10} = 3.90$. There was a significant effect of prediction type, $\chi^2(1) = 13.89$, $p < .001$, with strong support for the alternative hypothesis, BF_{10}

$= 1.23 \times 10^4$, showing more intrusions when Day 1 endings were predicted during Day 2 viewing. Although the interaction was not significant, $\chi^2(1) = 3.32$, $p = .07$, there was strong evidence for the alternative hypothesis, $BF_{10} = 10.02$. The finding of more intrusions following memory-based predictions when change was not remembered replicates Experiment 1 in showing the cost of prediction errors in the absence of recollection-based retrieval. The similar, but numerically smaller, difference when change was remembered was unique. We cannot explain this, especially since there was ambiguity in the evidence for the interaction effect between the two analytic approaches. Collectively, these conditional results replicate Experiment 1 in showing improved and impaired memory for Day 2 activities associated with memory-based predictions that depended on change recollection.

Discussion

The results of Experiment 2 replicated and extended Experiment 1. Changing a central activity feature was again associated with enhanced Day 2 recall when change was recollected and impaired recall when change was not recollected. The memory impairment associated with failed recollection also resulted in more Day 1 intrusions. Comparing across experiments, the instructions in Experiment 2 to make memory-based predictions for Day 2 endings appeared to improve Day 2 recall for repeated and changed endings. But this could also have reflected participants being sampled from different universities. Regardless, finding that a higher rate of memory-based predictions was associated with better subsequent memory emphasizes the benefits of prior-event retrieval on memory for all activity features. As in Experiment 1, conditional analyses showing that prediction errors were associated with both improved and impaired updating depending on change recollection provided provisional correlational evidence of a role for prediction error in updating. However, as in Experiment 1, there was no clear association between pure prediction errors and memory for novel activities, an association that would be expected if prediction error leads to memory updating. Collectively, these results clearly show that prior-event retrievals benefitted event memory updating and leave open the possibility that prediction error played a role, but do not conclusively establish such a role.

General discussion

In two experiments, we tested two proposed mechanisms of successful event change detection and memory updating: prior-event retrieval and prediction error. We did this by incorporating overt retrieval and prediction measures during event encoding into the everyday changes paradigm. This allowed us to assay retrieval success, control the activities

associated with prediction error, and examine the contributions of event retrieval and prediction error to memory for changed event features. We replicated earlier findings showing proactive facilitation in event memory retrieval when change was recollected and proactive interference when change was not recollected (Wahlheim & Zacks, 2019). Both experiments implicated a role for prior-event retrieval in enhanced recall of changed event features. Across experiments, there was inconsistent evidence for a direct association between prediction error and updating, as it was inconclusive in Experiment 1 and strong in Experiment 2. Both experiments showed clear evidence that prediction error was associated with more change recollection, which was associated with better updating, but prediction error was also associated with more intrusions when change was not recollected. This showed that prediction error effects on overall recall may be obscured by offsetting effects that depend on recollection-based retrieval. Below, we consider the theoretical implications of these findings for mechanisms of event memory updating.

Prior-event retrieval and event memory updating

Research in the episodic memory updating literature has implicated a role for retrieval of prior memories in learning and has demonstrated its benefits on subsequent memory (see Hintzman, 2011). This retrieval process also plays a central role during the specific encoding of changed features in effective memory updating. The MFC framework and EMRC propose that this retrieval process is necessary for past and present event features to coexist in consciousness, therefore enabling encoding of configural representations (Jacoby et al., 2015; Wahlheim & Jacoby, 2013; Wahlheim & Zacks, 2019). Consistent with this view, neuroimaging studies have shown that reactivation of prior memories while encoding competing memories is associated with interference reduction and improved memory updating (e.g., Chanales et al., 2019; Koen & Rugg, 2016; Kuhl et al., 2010; Stawarczyk et al., 2020).

In the present study, we measured prior event (Day 1) retrieval during overt predictions made while viewing Day 2 movies. We assumed that predictions based on memory for Day 1 features indicated successful prior-event retrieval that guided Day 2 predictions. In Experiment 1, we instructed participants to predict what they thought would happen without suggesting that they should make their predictions based on Day 1 memories. Allowing participants to choose their prediction strategy produced individual variation in memory-based predictions that was associated with subsequent memory. Specifically, participants who made more memory-based predictions showed higher subsequent memory for changed Day 2 features. In Experiment 2, when we instructed participants to use a memory-based prediction strategy, the overall rate of memory-based predictions was higher than in

Experiment 1. When participants were explicitly instructed to make memory-based predictions we found that, within-participants, successful memory-based predictions were predictive of subsequent memory of Day 2 event features. This suggests that encouraging participants to make memory-based predictions increased their frequency of retrieval, and that this facilitated memory. Of course, one must consider the caveat that this between-experiment difference could have reflected differences in the population from which the samples were drawn, as the experiments were conducted at different universities.

More broadly, the associations observed between memory-based predictions and Day 2 recall are consistent with previous findings showing that successful retrieval of related event features during encoding is associated with better memory accuracy for changed features (e.g., Wahlheim & Jacoby, 2013). Taken with the finding that Day 2 recall for changed features was also higher when change was recollected, these results suggest that memory-based predictions, and successful retrieval, facilitated the detection of changed features. However, we could not measure change detection directly here. One possibility is that memory-based prediction enabled the formation of integrated memory representations.

Prediction error in event comprehension

Behavioral and neurophysiological data suggest that prediction errors can trigger memory updating. For example, Kim et al. (2014) presented sequences of pictures, with repeating contingencies that could introduce predictions. Prediction was assayed using multivariate pattern fMRI, and it was found that prediction errors were associated with memory updating. Bein et al. (2019) used deliberate retrieval of picture memories to induce predictions that then led to errors when changed pictures were presented. They found that this shifted hippocampal activity from a retrieval-related state to an encoding-related state. Similarly, Sinclair et al. (2020) used a partial-reminding procedure with videos to induce prediction errors and found that such prediction errors both shifted hippocampal states and promoted memory updating (see also Sinclair & Barense, 2019). Such results are consistent with EMRC, as described in the Introduction (Wahlheim & Zacks, 2019).

However, across the present experiments, we found inconclusive evidence for a direct association between prediction error and memory updating. This could reflect the all-or-none nature of the memory-based prediction measure. Previous research has demonstrated that the strength of reactivation driving a prediction is important for modulating memory updating; weak or strong reactivation is less effective than intermediate reactivation (Bein et al., 2019; Kim et al., 2014; Norman et al., 2007; Sinclair & Barense, 2019). This could also reflect that the self-reported prediction measure used here was

insensitive to effects of prediction error on memory updating. The task we used here halted participants' event processing and forced them to make an explicit prediction by choosing one of two options. The video playback had been stopped, removing the participant from the context of the narrative, and a conscious choice had to be made. Although previous research has successfully used this measure to characterize the role of prediction error in event segmentation (e.g., Zacks et al., 2011), this task may not always be a valid indicator of ongoing prediction error during event comprehension.

Other studies have sought to assess prediction error by using indirect cognitive neuroscience techniques rather than more direct behavioral measures. For example, previous research has conceptualized and measured prediction error through hippocampal activity indicated by functional magnetic resonance imaging (fMRI; Shohamy & Wagner, 2008), dopaminergic activity measured by neural electrodes (Bayer & Glimcher, 2005), and through dopaminergic activity in the midbrain and striatal regions, again measured by fMRI (Zacks et al., 2011). Although such techniques can be leveraged to examine prediction error mediated learning, they are not always accessible. This indicates the need for more accessible behavioral measures that are sensitive to prediction error signals. One caveat of this approach, as shown in the present study, is that behavioral measures as indirect assays will vary in their sensitivity. The prediction-contingent paradigm used here represents an indirect measure that may not have been sufficiently sensitive to the kind of predictive processing that precedes the encoding of event changes. Ultimately, any direct assay of prediction error may not be effective as it would necessarily involve interrupting ongoing encoding, necessitating the call for indirect measures, whether behavioral or physiological.

Eye tracking is another indirect behavioral measure used to examine predictive processing during naturalistic event comprehension. This measure shows promise as a sensitive tool for detecting prediction error signals. For example, Eisenberg et al. (2018) demonstrated predictive looking during narrative activities with eye tracking by showing that viewers looked at objects just before actors contacted those objects. Eye tracking is an effective and unobtrusive way of measuring predictions during ongoing processing, without halting participants' engagement. Future studies may benefit from further identification of noninvasive assays of prediction error, such as eye tracking, during ongoing comprehension.

Conclusion

The present study tested theoretical mechanisms proposed to underlie the encoding and memory for naturalistic event changes. In two experiments, making predictions based on memory for a previous related event was associated with successfully encoding changes in a new event. This is consistent with models proposing that spontaneous retrieval during encoding facilitates that new encoding. Such models generally further predict that when things change, retrieval during encoding leads to prediction errors, which in turn drive memory updating. The current results did not provide strong evidence for or against this proposal. One possibility is that indirect assays may provide a more sensitive test of the hypothesis that prediction error drives memory updating for naturalistic events. We are currently exploring this possibility.

Instructions

Appendix A

Day 1 movie

In this part of the experiment, you will watch an actor perform a series of everyday activities throughout the course of a day. Your task will be to pay careful attention to her actions.

Day 2 movie

In this part of the experiment, you will watch an actor perform a series of everyday activities throughout the course of another day that occurred later in the week. These activities will vary in their relationship to those she performed on Day 1. She will repeat some activities in exactly the same manner as Day 1; she will perform some new activities that she did not perform on Day 1; and she will perform some activities in a manner similar to Day 1, but a critical feature of those activities will change. You should pay careful attention to her actions and note the relationship between these actions on Day 2 and the actions from Day 1.

Importantly, we are also interested in your ability to predict her future actions. To measure this, we will stop the video at various points and ask you to make an explicit prediction about which of two actions she will perform within the next few seconds. When the video stops, two images will appear, depicting the actor performing two different actions. Your task will be to choose the action that she is more likely to perform.

To do this, press the "Q" key to choose the image on the left side of the screen or press the "P" key to choose the image on the right side of the screen. After you have chosen an action, a screen will appear with the message "Imagine Actions" for 5 s. Take this time to imagine how the upcoming sequence of

actions will unfold. The video will resume after the 5 s have elapsed. Please pay attention to the actor's subsequent actions and compare them with the actions that you imagined.

Test

In this part of the experiment, you will be asked to recall details from the activities you viewed from Day 2. For each question, you will first be asked to type a response in a box below. After you have typed your response, press ENTER.

Here, we are also interested in your ability to remember which activities changed from Day 1 to Day 2. When you watched the actor perform activities on both days, she performed some activities in the exact same manner on both days, she performed other activities only on the second day, and she changed how she performed another set of activities from the first to second day.

After recalling details from each activity, you will be asked whether certain actions associated with an activity changed from Day 1 to Day 2. For example, the actor could have cleaned the counter with a paper towel on Day 1 and a washcloth on Day 2 (you will see an example of this change shortly). When you are asked whether the activity changed between days, you will press "1" to respond "Yes" and "2" to respond "No." When you respond "Yes" that you remember the activity changing from Day 1 to Day 2, you will be asked to report how the activity was accomplished on Day 1. Please type your response in the box and press ENTER.

Appendix B

Session 1

The purpose of this study is to investigate how people understand and remember everyday events. In this experiment, you'll first watch movies of an actor performing daily life activities, and afterwards we will ask you to answer questions about these movies. This experiment will take place in three sessions, with a week between each session, at the same time and place. Each session will last approximately 1 hour.

In this part of the experiment, you will watch an actor perform a series of activities throughout the course of a normal day. Pay close attention to the activities and try to commit them to memory. Press Space to continue and watch two examples of activities.

This is the type of activity you'll see during the real task. The first movie will take about 20 minutes, and we will call it "Day 1."

Session 2

In this part of the experiment, you will watch the same actor perform another series of everyday activities throughout the course of a second day. We'll refer to this as DAY 2 to distinguish it from the DAY 1 you already saw. You will first watch the beginning of each activity, and then it will be stopped and you will be asked to select what you think will happen next by pressing the number in parentheses that corresponds with the image of your choice, based on your memory of what happened on DAY 1. Then you will watch the remainder of the activity on DAY 2. Please note that some activities will have the same endings as DAY 1, and some will be different. Still pay close attention and try to commit the activities to memory. You will first complete two practice examples. Press Space to continue.

On both days, the actor placed her phone in the exact same spot on the piano, but she plugged in an e-reader to charge on Day 1, and on Day 2 it was a tablet. This is the kind of change that you will see during the real task. It is important to note these changes in detail.

Session 3

You will be presented with the beginning of one of the activities that you viewed previously. Remember the movie from the first session is DAY 1 and the movie from the second session is DAY 2. Once the clip has ended, you will be asked to recall what happened on DAY 2. You will type in your response, and then press Enter. After responding with the ending to DAY 2, you will be asked whether the ending changed from DAY 1 to DAY 2. Make your selection by pressing the key in parentheses that corresponds to your answer. If you select "changed," you will then be asked to report what happened on DAY 1. You will first be shown the two example clips as practice. Press Space to continue.

It's crucial that you answer with the Day 2 ending, and then, if it differed from Day 1, you will report the Day 1 ending afterwards. Make sure to respond with as much detail as possible.

References

Anderson, M. C., & Neely, J. H. (1996). Interference and inhibition in memory retrieval. In E. L. Bjork & R. A. Bjork (Eds.), *Handbook of perception and cognition Memory* (2nd ed., pp. 237–313). Academic Press.

Ausubel, D. P., Robbins, L. C., & Blake, E., Jr. (1957). Retroactive inhibition and facilitation in the learning of school materials. *Journal of Educational Psychology, 48*(6), 334–343. https://doi.org/10.1037/h0043524

Barnes, J. M., & Underwood, B. J. (1959). "Fate" of first-list associations in transfer theory. *Journal of Experimental Psychology, 58*, 97–105. https://doi.org/10.1037/h0047507

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48. https://cran.r-project.org/web/packages/lme4/vignettes/lmer.pdf

Bayer, H., & Glimcher, P. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron, 47,* 129–141. https://doi.org/10.1016/j.neuron.2005.05.020

Bein, O., Duncan, K., & Davachi, L. (2019). *Mnemonic prediction errors bias hippocampal states*. bioRxiv. https://doi.org/10.1101/740563

Bruce, D., & Weaver, G. E. (1973). Retroactive facilitation in short-term retention of minimally learned paired associates. *Journal of Experimental Psychology, 100*, 9–17. https://doi.org/10.1037/h0035488

Bürkner, P. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal, 10*, 395–411.

Chanales, A. J. H., Dudukovic, N. M., Richter, F. R., & Kuhl, B. A. (2019). Interference between overlapping memories is predicted by neural states during learning. *Nature Communications, 10*(1), Article 5363. https://doi.org/10.1038/s41467-019-13377-x

Eisenberg, M., Zacks, J. M., & Flores, S. (2018). Dynamic prediction during perception of everyday events. *Cognitive Research: Principles and Implications, 3,* 53. https://doi.org/10.1186/s41235-018-0146-z

Faul, F., Erdfelder, E., Buchner, A., Buchner, A., & Lang, A. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Fox, J., & Weisberg, S. (2011). *An {R} companion to applied regression* (2nd ed.). SAGE.

Hintzman, D. L. (2004). Judgment of frequency vs. recognition confidence: Repetition and recursive reminding. *Memory & Cognition, 32*, 336–350. https://doi.org/10.3758/BF03196863

Hintzman, D.L. (2010). How does repetition affect memory? Evidence from judgments of recency. *Memory & Cognition, 38*, 102–115. https://doi.org/10.3758/MC.38.1.102

Hintzman, D. L. (2011)Research strategy in the study of memory: Fads, fallacies, and the search for the "coordinates of truth" *Perspectives on Psychological Science, 6*, 253–271. https://doi.org/10.1177/1745691611406924

Jacoby, L. L., Wahlheim, C. N., & Kelley, C. M. (2015). Memory consequences of looking back to notice change: Retroactive and proactive facilitation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*, 1282–1297. https://doi.org/10.1037/xlm0000123

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford University Press.

Kass, R., & Raftery, A. (1995). Bayes factors. *Journal of the American Statistical Association, 90,* 773–795.

Kim, G., Lewis-Peacock, J., Norman, K., & Turk-Browne N. (2014). Pruning of memories by context-based prediction error. *Proceedings of the National Academy of Sciences of the United States of America, 111,* 8997–9002. https://doi.org/10.1073/pnas.1319438111

Koen, J. D., & Rugg, M. D. (2016). Memory reactivation predicts resistance to retroactive interference: Evidence from multivariate classification and pattern similarity analyses. *The Journal of Neuroscience, 36*(15), 4389–4399. https://doi.org/10.1523/JNEUROSCI.4099-15.2016

Kuhl, B. A., Shah, A. T., DuBrow, S., & Wagner, A. D. (2010). Resistance to forgetting associated with hippocampus-mediated reactivation during new learning. *Nature Neuroscience, 13*, 501–506. https://doi.org/10.1038/nn.2498

Lenth, R. V. (2019). Estimate marginal means, aka least-squares means: The R Package emmeans [Computer software]. https://CRAN.R-project.org/package=emmeans

McGeoch, J. A., & McDonald, W. T. (1931). Meaningful relation and retroactive inhibition. *The American Journal of Psychology, 43,* 579–588. https://doi.org/10.2307/1415159

Norman, K. A., Newman, E. L., & Detre, G. (2007). A neural network model of retrieval-induced forgetting. *Psychological Review, 114*(4), 887–953. https://doi.org/10.1037/0033-295X.114.4.887

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman,E., Lindeløv, J. (2019). PsychoPy2: experiments in behavior made easy. *Behavior Research Methods, 51,* 195–203. https://doi.org/10.3758/s13428-018-01193-y

Psychology Software Tools, Inc. [E-Prime 3.0]. (2016). Retrieved from http://www.pstnet.com.

R Core Team. (2013). R: A language and environment for statistical computing [Computer software]. R Foundation for Statistical Computing. http://www.R-project.org/

Radvansky, G. A., (2005). Situation models, propositions, and the fan effect. *Psychonomic Bulletin & Review, 12*, 478–483. https://doi.org/10.3758/BF03193791

Robbins, D., & Bray, J. F. (1974). The spacing effect and the A-B, A-C paradigm: Evidence for retroactive facilitation. *Journal of Experimental Psychology, 103*, 420–425. https://doi.org/10.1037/h0037203

Shohamy, D., & Wagner, A. (2008). Integrating memories in the human brain: Hippocampal-midbrain encoding of overlapping events. *Neuron, 60,* 378–389. https://doi.org/10.1016/j.neuron.2008.09.023

Sinclair, A., & Barense, M. (2019). Prediction error and memory reactivation: How incomplete reminders drive reconsolidation. *Trends in Neuroscience, 42,* 727–739. https://doi.org/10.1016/j.tins.2019.08.00

Sinclair, A., Manalili, G., Brunec, I., Adcock, R., & Barense, M. (2020). *Prediction errors during naturalistic events modulate hippocampal representations and drive episodic memory updating*. bioRxiv. https://doi.org/10.1101/2020.09.29.319418

Stawarczyk, D., Wahlheim, C., Etzel, J., Snyder, A., & Zacks, J. M. (2020). Aging and the encoding of changes in events: The role of neural activity pattern reinstatement. *Proceedings of the National Academy of Sciences of the United States of America, 117*(47) 29346–29353. https://doi.org/10.1073/pnas.1918063117

Underwood, B. J. (1957). Interference and forgetting. *Psychological Review, 64*(1), 49–60. https://doi.org/10.1037/h0044616

Wahlheim, C. N. (2014). Proactive effects of memory in young and older adults: The role of change recollection. *Memory & Cognition, 42*, 950–964. https://doi.org/10.3758/s13421-014-0411-4

Wahlheim, C. N. (2015). Testing can counteract interference by integrating competing information. *Memory & Cognition, 43*, 27–38. https://doi.org/10.3758/s13421-014-0455-5

Wahlheim, C. N., & Jacoby, L. L. (2013). Remembering change: The critical role of recursive remindings in proactive effects of memory. *Memory & Cognition, 41*, 1–15. https://doi.org/10.3758/s13421-012-0246-9

Wahlheim, C. N., Smith, W. G., & Delaney, P. F. (2019). Reminders can enhance or impair episodic memory updating: A memory-for-change perspective. *Memory, 27*(6), 849–867. https://doi.org/10.1080/09658211.2019.1582677

Wahlheim, C. N., & Zacks, J. M. (2019). Memory guides the processing of event changes for older and younger adults. *Journal of Experimental Psychology: General, 148*(1), 30–50. https://doi.org/10.1037/xge0000458

Zacks, J. M., Kurby, C. A., Eisenberg, M. L., & Haroutunian, N. (2011). Prediction error associated with the perceptual segmentation of naturalistic events. *Journal of Cognitive Neuroscience, 23*(12), 4057–4066. https://doi.org/10.1162/jocn_a_00078

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin, 133*(2), 273–293. https://doi.org/10.1037/0033-2909.133.2.273

Open practices statement

The materials, data, and analysis scripts for both experiments are available (https://osf.io/7y5uv/).