Understanding Everyday Events: Predictive Looking Errors Drive Memory Updating

Journal:	*Psychological Science*
Manuscript ID	PSCI-21-0054.R1
Manuscript Type:	Research Article
Keywords:	Cognition(s), Eye Movements, Memory, Prediction, Visual Attention

SCHOLARONE™
Manuscripts

Understanding Everyday Events:

Predictive Looking Errors Drive Memory Updating

Christopher N. Wahlheim[1], Michelle L. Eisenberg[2], David Stawarczyk[2,3], and Jeffrey M. Zacks[2]

Department of Psychology, University of North Carolina at Greensboro[1],

Department of Psychological and Brain Sciences, Washington University in St. Louis[2]

Department of Psychology, Psychology and Neuroscience of Cognition Research Unit,

University of Liège[3]

Correspondence concerning this article should be addressed to Christopher N. Wahlheim, Department of Psychology, University of North Carolina at Greensboro, 296 Eberhart Building, PO Box 26170, Greensboro, NC 27402, Email: cnwahlhe@uncg.edu.

For Review Only

Running head: PREDICTIVE LOOKING AND MEMORY UPDATING

Article type: Research Article

Submitted: 1/14/21

Revision accepted: 9/9/21

Abstract

Memory-guided predictions can improve event comprehension by guiding attention and the eyes to the location where an actor is about to perform an action. But when events change, viewers may experience predictive looking errors and need to update their memories. In two experiments (*N*s = 38 and 111), we examined the consequences of mnemonic predictive looking errors for comprehending and remembering event changes. University students watched movies of everyday activities with actions that repeated exactly and actions that repeated with changed features—for example, an actor reached for a paper towel on one occasion and a dish towel on the next. Memory guidance led to predictive looking errors that were associated with *better* memory for subsequently changed event features. These results indicate that retrieving recent event features can guide predictions during unfolding events, and that error signals derived from mismatches between mnemonic predictions and actual events contribute to new learning.

Keywords: action observation, event cognition, memory updating, mnemonic prediction error, predictive looking

For Review Only

Statement of Relevance

The present research employs a rigorous behavioral and physiological approach to identify the mechanisms involved in updating memories for naturalistic actions. This research is of general interest because it combines theory and methods across areas of cognitive science and cognitive neuroscience in a highly naturalistic paradigm. Our theoretically motivated approach leverages methodological innovations from the action observation, memory updating, event cognition, and visual attention literatures. An attractive feature of the paradigm is that it goes beyond single trial action observation by depicting an actor performing sequences of actions across physical locations (e.g., various rooms in her house). The results implicate a critical role for mnemonic predictions in updating memory for changed event features. These findings will improve our understanding of how people remember related content from separate occasions and the role that expectations based on earlier observations play in how well new information is learned.

For Review Only

Understanding Everyday Events:

Predictive Looking Errors Drive Memory Updating

Predicting the actions of others is essential for skilled social interactions. For example, to cooperate in preparing dinner, one may need to anticipate when their partner will use the sink to wash vegetables. Prediction can improve action comprehension and facilitate one's own action planning (Gredebäck, 2018; Gredebäck & Falck-Ytter, 2015). Anticipating the future based on the past is adaptive because people often repeat behaviors. But when situations change, people behave differently. If the partner in the example above purchased pre-washed vegetables on a later occasion, she may bypass washing them in the sink, leading the observer to experience a mnemonic prediction error. Viewers' comprehension of such unexpected action changes may require registering such mismatches between predicted and actual events.

Action observation studies show that adult viewers make online predictions about future actions, especially actions they have experienced. Viewers look ahead to contacted objects (Eisenberg et al., 2018; Flanagan & Johansson, 2003; Hayhoe et al., 2003; Land & McLeod, 2000) and learn to predict object trajectories from repeated sequences (Barnes et al., 2005; Diaz et al., 2013). Infants learn to predict manual actions that are repeatedly performed (Cannon et al., 2012; Falck-Ytter et al., 2006), show positive correlations between predictive looking and their experience performing the actions (Cannon et al., 2012; Gredebäck et al., 2018; Melzer et al., 2012), and look earlier when observing recently learned actions (Gerson & Woodward, 2014). Thus, memory for past actions can guide action prediction and comprehension.

When actions change across occasions, mnemonic predictions trigger surprise responses that may stimulate new learning and update predictions for future actions (Gredebäck et al., 2018). Infant eye tracking studies show longer looking times to events that violate expectations

For Review Only

about world knowledge (Gredebäck et al., 2018; Juvrud et al., 2019; Stahl & Feigenson, 2015). Infants and children also show greater pupil dilation when actions end unexpectedly (Gredebäck et al., 2018; Juvrud et al., 2019). Such prediction errors increase exploration and new learning. For example, infants who watched a ball unexpectedly roll through a wall later explored the ball more than infants who had seen a ball stop. When infants were taught that the ball squeaked, infants whose expectations has been violated better remembered the new property of the ball (Stahl & Feigenson, 2015). These findings show that when an action violated world knowledge, infants actively updated their understanding of the object's properties.

Conditions associated with prediction errors are also associated with memory updating. When learned image sequences later include unexpected images that trigger prediction errors, memory is reduced for expected but no longer relevant images (Kim et al., 2014). In addition, after people learn to associate consistently valanced words with a scene category, pairing similar scenes with oppositely valanced words leads to better memory (Greve et al., 2017). Similarly, switching sequential contingencies such that symbol cues are first followed by objects from one semantic category and then another enhances recollection for unexpected objects (Kafkas & Montaldi, 2018). Also, prediction errors based on memory for virtual scenes enhance subsequent memory, with the benefit increasing with the number of changed features (Bein et al., 2020). When viewing action sequences, interrupting repetitions before expected outcomes distorts memory for prior events, indicating a form of error-driven memory updating (Sinclair & Barense, 2018). Relatedly, during basketball viewing, prediction errors and subsequent memory are associated with increased pupil size and cortical activity shifts (Antony et al., 2020).

When viewers see an actor perform everyday activities that change over time, prediction errors may stimulate better encoding. In one series of experiments, viewers watched an actor

For Review Only

perform action sequences on two occasions with some actions changing (Wahlheim & Zacks, 2019). For example, while the actor initially retrieved a *bath towel* from the closet, she later opened the closet the same way but retrieved a *hand towel*. Memory for changed actions was better when changes were noticed and later recollected. Also, greater neural reinstatement of the original actions before viewing changed actions was associated with better subsequent memory (Stawarczyk et al., 2020). This suggests that anticipating repeated actions triggered updating for changes. Although neural reinstatement may be necessary for mnemonic predictions (Bein et al., 2020), this measure alone does definitively index predictive processing.

In the current study, we assayed mnemonic prediction error in this *everyday changes* paradigm using anticipatory eye movements as a converging measure. In the first session, participants watched two movies of an actor performing everyday activity sequences depicting two fictional "days" in her life (Day 1 and Day 2 movies). *Changed activities* started with the same actions in both movies but ended with different actions in each movie, whereas *repeated activities* started and ended with the same action in both movies. Experiment 2 also included *control activities* with actions that were only performed in the second movie. Eye movements were recorded while participants watched the movies. In the second session, one week later, participants recalled Day 2 activity features (i.e., contacted objects), indicated whether the contacted objects had changed between movies, and if so, recalled the Day 1 features.

Changed Day 2 features are better remembered when participants can also remember that those features had changed and the original Day 1 features (Garlitch & Wahlheim, 2021; Hermann et al., 2021; Stawarczyk et al., 2020; Wahlheim & Zacks, 2019). According to *Event Memory Retrieval and Comparison Theory* (EMRC; Wahlheim & Zacks, 2019), the memory

For Review Only

benefit for changed Day 2 features associated with recall of Day 1 features occurs partly because mnemonic prediction errors stimulate encoding of new features. EMRC proposes a causal cascade in which viewers may (a) retrieve the activity's previous ending, (b) use the remembered action to predict the Day 2 ending, (c) experience a mnemonic prediction error when they view an unexpected action change, and (d) update their memory in response to that error. This view is consistent with findings of error-driven associative learning (Rescorla & Wagner, 1972).

The present two experiments tested the hypothesis that while viewing the beginnings of changed activities, predictive looking based on memory for Day 1 endings should be associated with better subsequent memory for changes, recall of Day 1 features, and memory for the changed Day 2 features. We tested this against the hypothesis that predictive processing is not necessary for changed features to be better remembered when they are recollected as such. This could occur when viewers retrieve Day 1 activities *after* viewing changed Day 2 activity endings because postdictive processes (cf. Neely et al., 1989) may be used to compare activities and improve the encoding changed activity endings.

Method

Participants

For Review Only

In Experiment 1, we set the sample size based on previous studies using the everyday changes paradigm to measure memory effects (Wahlheim & Zacks, 2019; $N = 36$) because no relevant previous data were available to support a power analysis. We recruited 43 participants and excluded five either because we could not track their eyes (four) or due to attrition (one). The final sample included 38 participants (13 women) ages 18-27 ($M = 20.37$, $SD = 2.17$). In Experiment 2, we ran bootstrapping power analyses using custom codes in R software (R Core Team, 2020) for each of the primary hypotheses using the data from Experiment 1. We sampled Experiment 1 data with replacement to obtain varying numbers of simulated participants. We then ran mixed-effects models for the results of interest and determined if they were significant. We ran 1000 iterations for the recall models and 500 iterations for the eye tracking models. We determined the proportion of significant results, which provided the estimates for power. The results indicated that a sample of 90 participants would be enough to achieve 80% power. Consequently, we recruited 111 participants, anticipating that some would have to be excluded. We excluded 13 participants either because we could not track their eyes (12) or the equipment failed (1). The final sample included 98 participants (56 women) ages 18-25 ($M = 19.61$, $SD = 1.39$). All participants received course credit or $10 per hour.

For Review Only

Materials and Design

Both experiments included the same two movies showing a woman actor performing sequences of everyday activities in or around her home and workplace on two fictional “days” in her life. There were two versions (A and B) of each activity that included the same peripheral features and a changed central feature that always showed her manipulating an object (e.g., inserting a key into a door lock). Figure 1A shows key moments from two versions of an example activity where the actor unlocked her front door to enter her home. In Version A, she

Manuscript under review for Psychological Science

unlocked the *doorknob*, whereas in Version B, she unlocked the *deadbolt*. The different central features in each version are displayed at the point of contact ("Contact").

Experiment 1 used a 2 (Activity Type: repeated vs. changed) within-subjects design to manipulate the relationship between activities in each movie. Each movie included 44 critical activities (22 per condition) and 20 filler activities (10 per condition) that served to maintain narrative continuity (64 activities total). We counterbalanced the assignment of activity versions across activity types and movies in the following way. First, we created a Day 1 movie that included all critical activities assigned to either Version A or B. No more than three activities of the same version appeared sequentially. Then, we created another Day 1 movie, switching all the activity versions (i.e., Version A activities became Version B activities and vice versa). Finally, we created two Day 2 movies—one that included all Version A activities and one that included all Version B activities. The combinations of the Day 1 and Day 2 movies produced four experimental formats.

Experiment 2 included the same repeated and control conditions as in Experiment 1, but also included a Control condition with activities that only appeared in the Day 2 movie. Thus, it used a 3 (Activity Type: repeated vs. control vs. changed) within-subjects design. We assigned twice the number of activities to the changed condition than the other conditions to focus power on changed activities, which were of primary theoretical interest. The Day 1 movies included 33 critical activities (22 changed and 11 repeated) and 20 filler activities (10 changed and 10 repeated) for a total of 53 activities. The Day 2 movies included 44 critical activities (22 changed, 11 repeated, and 11 control) and 20 filler activities (10 changed and 10 repeated) for a total of 64 activities. The activities were counterbalanced in the same way as Experiment 1.

For Review Only

Both experiments used the same cued recall test items. The test included 64 questions probing memory for all activities from the Day 2 movie. Cues asked about the central features that could have changed. For example, the cue for the door-unlocking example in Figure 1 asked about the central *lock* feature (i.e., "Which lock did the actor unlock to enter her home?").

Procedure

Figure 2 displays a schematic of the two experimental sessions, which were separated by approximately one week (Experiment 1: M = 7.07 days, SD = 0.81 days, *range* = 7-11 days; Experiment 2: M = 7.07, SD = 0.50, *range* = 6-10 days). Both experiments used the same procedure with one exception noted below. The exact instructions shown to participants are available in the Supplemental Material. Movies were shown on a 19 in (74 cm) monitor (1440 × 900 resolution) at a 1280 × 720 aspect ratio using SR Research Experiment Builder software (SR Research Ltd., Mississauga, ON, Canada). Gaze location was recorded from the right eye using an infrared pupil-corneal eye tracker (EyeLink 1000; SR Research Ltd., Mississauga, ON, Canada) that sampled at 1000 Hz. Participants placed their head against a chin/forehead rest to minimize head motion. The camera was positioned 52 cm from the top of the rest. The viewing distance was 58 cm from the rest with a viewing angle of 38.6°.

During Session 1, participants were told that they would watch movies of an actor performing everyday activities and to pay attention to those activities. Participants watched two different movies from ostensibly separate "days" in the actor's life. As described above, half of the Day 2 activities included changed endings. While watching each movie, participants took short breaks between scenes showing morning, work, afternoon, and evening activities.

For Review Only

During Session 2, participants completed the cued recall test presented via E-Prime 2 software (Psychology Software Tools, Pittsburg, PA) in another room. They were told that they would recall activity features and indicate which features had changed between movies. Before the test, participants viewed two example clips of a hair styling activity with a changed feature. The actor styled her hair first with a comb and then with a brush. Test items then appeared individually (for a schematic of the test trial structure, see Figure 2B). On each trial, participants first recalled the Day 2 feature by typing their response. Then they indicated if the feature had changed by responding "yes" or "no" with the "1" or "2" key (Experiment 1) or by classifying the activity as repeated, changed, or only shown on Day 2, with the "1", "2", or "3" key (Experiment 2). When participants indicated that an activity had changed ("yes" or "changed"), they attempted to recall the Day 1 feature by typing their response. When participants indicated than an activity had not changed, the program advanced to the next trial.

Statistical Analysis

All analyses were conducted using R software. We fitted linear and logistic mixed effects models including experimental variables as fixed effects and subjects and activities (items) as random effects using functions from the *lme4* package (Bates et al., 2015). We tested for significant effects of predictor variables using the Wald test in the *Anova* function from the *car* package (Fox & Weisberg, 2019). We conducted pairwise comparisons using the *emmeans* function from the *emmeans* package (Lenth, 2020) with the Tukey method controlling for multiple comparisons. All statistical tests were two-sided. The level for significance was set at $\alpha = .05$. The probabilities and confidence intervals below were estimated from these models.

Results

For Review Only

Each activity included a pair of critical objects that the actor might manipulate, such as the doorknob and deadbolt in Figure 1A. For repeated activities, the actor manipulated the same *target* object on both days—for example, she could unlock the deadbolt on both days (Figure 1B). In that case the *alternate* object on both days would be the doorknob. For changed activities, the target object in the Day 1 movie became the alternate object in the Day 2 movie, and vice versa (Figure 1C). Both repeated and changed activities appeared in both experiments. Control activities that the actor only performed in the Day 2 movie also appeared in Experiment 2. The primary eye tracking measure of interest was the proportion of time looking to the target and alternate objects while watching each activity. We report results from analyses comparing these proportions across activity types and movies from both experiments together. We also conducted exploratory analyses of pupil areas that we report in the Supplemental Material.

Proportions of Looks to Interest Areas

To characterize looking patterns as activities unfolded, the co-first authors first divided each activity into two intervals of interest (Figure 1A). The raters watched each version of each activity together and jointly decided about interval placement. The *pre-divergence* interval started when the target and alternate object were both visible and ended the moment before the actor started to move more towards the target than alternate object (e.g., the first moment her reach trajectory indicated that she was more likely to contact the deadbolt than doorknob). The *post-divergence* interval started the moment after this divergence point and ended when the actor contacted the target object. Total interest intervals ranged from 470 to 29,200 ms (M = 6434 ms, SD = 5032 ms). Pre-divergence intervals (*range* = 100 to 16,920 ms, M = 5212 ms, SD = 3597 ms) were significantly longer than post-divergence intervals (*range* = 140-13,250 ms, M = 1222 ms, SD = 2447 ms), $t(87) = 10.61$, $p < .001$.

For Review Only

Because the pre- and post- divergence intervals varied in length, we divided each interval into 10 isochronous bins for analysis (20 total bins). This allowed us to compare the time course of looking behavior in terms of the relative time elapsed from the onset of each activity to the divergence point and from the divergence point to object contact, controlling for differences in duration. To analyze looking patterns as activities unfolded, we included time bin (Bin) as a categorical predictor variable in the mixed-effects models. When testing hypotheses about the effects of prior viewing on predictive looking before and after divergence points, we used separate models that only included either the first 10 bins (pre-divergence) or second 10 bins (post-divergence). For each activity, we defined spatial regions of interest by drawing polygons around the target and alternate objects (Figures 1B & C). We then recorded, for each participant, whether they looked in one or both interest areas during each time bin. In the following analyses, we report looking proportions in each interest area aggregated across all activities. In addition to the analyses reported here, we examined the consistency of looking proportion differences across activities based on pre- and post-divergence interval lengths (see Supplementary Material).

When Viewing Activities Again, Viewers Look Predictively to Previous Targets of Action

Repeated Activities. For repeated activities, the actor manipulated the same object on both days, so mnemonic predictions based on Day 1 actions should lead to more looking to Day 2 target objects. This was what we found (Figure 3, top left quadrants in both panels). In both experiments, logistic mixed-effects models were used to estimate proportions of looking to each object of interest (Object) during both movies (Day) at each time bin (Bin). Models including Object, Day, and Bin as fixed effects (Table 1, top rows) indicated significant Object × Day interactions. Participants looked significantly more to target objects when watching the Day 2

For Review Only

than Day 1 movie, smallest z ratio = 9.39, $p < .001$. In contrast, they looked significantly less to alternate objects when watching the Day 2 than Day 1 movie, smallest z ratio = 2.03, $p = .043$.

The models in both experiments also indicated significant Object × Bin interactions showing the following patterns: In the pre-divergence intervals (time bins -9 to 0; from "Start" to "Divergence"), looking proportions were mostly consistent across adjacent bins. Pairwise comparisons only indicated significant increases for target objects in both experiments from bin -1 to 0, smallest z ratio = 4.83, $p < .001$. In the post-divergence intervals (time bins 1 to 10; from the bin after "Divergence" to "Contact), looking patterns differed for target and alternate objects. For target objects, looking proportions increased consistently until Bin 6 (Experiment 1) and Bin 7 (Experiment 2) and then began to asymptote. In contrast, for alternate objects in both experiments, there were overall decreases in looking proportions across the interest periods. Looking proportions were significantly greater in the first bin (1) than the last bin (10), smallest z ratio = 7.41, $p < .001$. However, the decrease in looking was sharpest across the three bins after the divergence point, as there were few significant differences beyond Bin 3.

There was also an Object × Day × Bin interaction in Experiment 2. To characterize this interaction, separate models with the same variables were fitted to data including only the time bins from pre- or post-divergence intervals. In the pre-divergence model (Table 2, top rows), an Object × Day interaction showed significantly greater proportions of looking to target objects in the Day 2 than Day 1 movie, z ratio = 4.76, $p < .001$, but no significant difference in the proportions of looking to alternate objects in each movie, z ratio = 1.19, $p = .234$. In the post-divergence interval, an Object × Day interaction indicated significantly greater proportions of looking to target objects in the Day 2 than Day 1 movie, z ratio = 8.55, $p < .001$, and to alternate

For Review Only

objects in the Day 1 than Day 2 movie, z ratio = 4.33, $p < .001$. These results show that viewing experience increased predictive looking to objects the actor would contact early in each activity.

Changed Activities. For changed activities, the actor manipulated one object in the Day 1 movie but then a different object in the Day 2 movie. The objects that were manipulated on Day 1 but not Day 2 are referred to as alternate objects for Day 2 (see Figure 1C). During the early part of a changed activity on Day 2, mnemonic predictions based on actions from Day 1 movies should increase looking to the alternate object, because that object was previously the action target. After the divergence point, as visual cues provide information that the activity will end differently, looking to the target should increase. Figure 3 (right quadrants in both panels) shows that this qualitative pattern was observed in both experiments.

For Experiment 1, a model with the same variables as in the looking proportion model for repeated activities (Table 1, bottom rows) indicated a significant Object × Day × Bin interaction. This interaction was characterized using separate models for the pre- and post-divergence intervals. In the pre-divergence interval (Table 2, bottom rows), a significant Object × Day interaction showed greater proportions of looking to both objects when viewers watched the Day 2 than Day 1 movie, with this difference being greater for alternate, z ratio = 6.01, $p < .001$, than target, z ratio = 1.99, $p = .047$, objects. These results suggest that memory for actions in Day 1 movies guided viewers' looking towards areas including objects that could be contacted in Day 2 movies, especially for areas including the earlier-contacted objects. In the post-divergence interval, a significant Object × Day interaction indicated greater proportions of looking to target objects in the Day 2 than Day 1 movie, z ratio = 10.00, $p < .001$, and to alternate objects in the Day 1 than Day 2 movie, z ratio = 2.37, $p = .018$. These results suggest that observed movement

For Review Only

trajectories that contradicted mnemonic predictions stimulated more looking to objects that had not been contacted in the Day 1 movie.

In both experiments, there were also significant Object × Bin interactions indicating the following patterns: In the pre-divergence intervals, looking proportions were mostly consistent across adjacent bins with some exceptions. For target objects in both experiments, there were significant increases from bin -1 to 0, smallest z ratio = 5.19, $p < .001$. For target objects in Experiment 2, the first bin (-9) was significantly lower than most other bins, smallest z ratio = 3.17, $p = .049$. For alternate objects in Experiment 2, there were significant increases in bins -8, -7, -1, and 0, smallest z ratio = 3.20, $p = .045$. In the post-divergence intervals, there were substantial differences in proportions of looking to target and alternate objects. For targets in both experiments, there were generally consistent increases in looking across adjacent bins, until Bin 7. In contrast, for alternate objects in both experiments, looking proportions decreased as activities unfolded. Looking proportions were significantly greater in the first bin (1) than last bin (10), smallest z ratio = 8.65, $p < .001$. However, the decrease in looking to alternate objects was sharpest in the four bins after the divergence point, as there were few significant differences beyond Bin 4.

Finally, in Experiment 2, there was a significant effect of Day, replicating the finding from Experiment 1 that proportions of looking to both objects were greater during Day 2 than Day 1 movies. However, in contrast to Experiment 1, neither the Object × Day nor the Object × Day × Bin interaction was significant. In sum, when watching an activity for the second time, viewers looked predictively to the target object that the actor had previously manipulated. Viewers did this starting early in the activity, before the two potential endings diverged. For

For Review Only

changed activities, this resulted in mnemonic predictive looking errors that were subsequently corrected as the actor's hand approached the changed target object.

Cued Recall Performance

When participants were asked about features of Day 2 activities on the cued recall test, most responses were correct reports of the object manipulation seen on Day 2 (Day 2 recalls: 57% [Experiment 1], 50% [Experiment 2]) or incorrect reports of the alternate object manipulation (alternate intrusions: 20% [Experiment 1], 21% [Experiment 2]). In the example changed activity shown in Figure 1C, reporting "she unlocked the deadbolt" would be a correct Day 2 recall, and reporting that "she unlocked the doorknob" would be an alternate intrusion. Note that alternate intrusions are intrusions from episodic memory only for changed activities; for repeated and control activities, such responses reflect false remembering of an action that was not performed but fit within the semantic context of the activity. The remaining responses included ambiguous responses (4% [Experiment 1], 9% [Experiment 2]) that were correct but did not identify a target or alternate features (e.g., "she unlocked the house"), and other errors such as recalling other activity features or omissions (19% [Experiment 1], 20% [Experiment 2]).

Event Changes Can Lead to Proactive Facilitation in Memory

Both experiments showed that repeating an activity led to overall better correct recall of central features and fewer alternate intrusions than did repeating an initial action sequence and including a changed ending (Figure 4, black points). In Experiment 2, we also replicated the somewhat surprising finding that repeating an activity with a changed feature can lead to *facilitation* rather than interference in overall correct recall (Wahlheim & Zacks, 2019): Recall of central features for changed activities was better than recall of such features for control activities seen only on Day 2 (bottom left panel). Separate models estimating correct Day 2 recall and

For Review Only

alternate intrusion probabilities including only Activity Type as a fixed effect (repeated and changed activities in Experiment 1; repeated, control, and changed activities in Experiment 2), indicated significant effects for Day 2 recall [Experiment 1: $\chi^2(1) = 6.59$, $p = .010$; Experiment 2: $\chi^2(2) = 60.32$, $p < .001$] and alternate intrusions [Experiment 1: $\chi^2(1) = 28.21$, $p < .001$; Experiment 2: $\chi^2(2) = 33.58$, $p < .001$]. For Experiment 2, all pairwise comparisons among the three levels of activity type were significantly different for Day 2 recall (smallest z ratio = 3.43, $p = .002$), whereas the only significant difference for alternate intrusions was between repeated and control activities, z ratio = 1.86, $p = .150$.

For Review Only

Recollecting that Activities had Changed is Associated with Memory Facilitation

To test whether successful updating was associated with recollecting that an activity had changed, we analyzed cued recall performance for changed activities conditionalized on classifications indicating if activities had changed, made just after each cued recall attempt. We created three response classification types for changed activities: *Change Recollected* were accurate classifications with recall of Day 1 features; *Change Remembered* were accurate classifications without correct recall of Day 1 features; and *Change Not Remembered* were inaccurate classifications. Change classification rates were comparable in both experiments: *Recollected* (Experiment 1 = 38%; Experiment 2 = 37%), *Remembered* (Experiment 1 = 19%; Experiment 2 = 18%); and *Not Remembered* (Experiment 1 = 43%; Experiment 2 = 45%).

Correct recall of changed Day 2 features varied depending on how participants classified changed activities at test (Figure 3, green, blue, and red points). In both experiments, models fitted to correct Day 2 recall including Classification as the fixed effect indicated significant effects, smallest $\chi^2(2) = 188.29$, $p < .001$. Pairwise comparisons indicated that recall was significantly greater when change was recollected (green points) than when it was remembered but not recollected (blue points) or not remembered (red points), smallest *z* ratio = 11.13, $p <$ .001, and that the latter two conditions (blue and red points) were not significantly different, largest *z* ratio = 0.84, $p = .678$. The results replicate findings showing that enhanced event memory updating for changed activities was associated with change recollection (Garlitch & Wahlheim, 2021; Hermann et al., 2021; Stawarczyk et al., 2020; Wahlheim & Zacks, 2019).

In both experiments, alternate intrusions were most likely when change was remembered but not recollected (blue points) and least likely when change was recollected (green points). The classification effect was significant in both experiments, smallest $\chi^2(2) = 90.96$, $p < .001$. All pairwise comparisons were significantly different, smallest *z* ratio = 3.54, $p = .001$.

Predictive Looking Errors are Associated with Better Event Memory Updating

The previous analyses established that viewers sometimes make mnemonic predictive looking errors when watching Day 2 activities unfold, and that watching a changed activity ending can lead to facilitation, rather than interference, in memory for changed activity features. We hypothesized that these two results are related, such that making a predictive looking error can induce memory updating (cf. Stahl & Feigenson, 2015). To test this hypothesis, we assessed proportions of looking while viewers watched changed activities on Day 2 back-sorted on whether changes were later recollected (Figure 5). We compared activities for which changes were recollected, i.e., that were correctly classified and for which the Day 1 activity feature was recalled (Recollected), to activities for which change was not recollected (Not Recollected). The latter category included activities for which Day 1 features were not recalled, regardless of whether activities were classified as having changed, because recall of Day 2 features in those cells was not significantly different (see Figure 4, left panels, blue and red points).

For Review Only

For changed activities in the Day 2 movie, mnemonic predictive looking errors are looks to alternate objects that the actor had manipulated in the Day 1 movie. Consistent with our hypothesis, Figure 5 (left panels) shows that in both experiments, proportions of looking to alternate objects for changed activities were greater during activities for which viewers subsequently recollected that features had changed (green lines). This was especially true when viewers watched the actor repeat actions from the Day 1 movie in the pre-divergence intervals.

In both experiments, models fitted to looks to alternate objects with Change Recollection (Recollected vs. Not Recollected [vs. Control in Experiment 2 only]) and Bin as fixed effects showed significant main effects (Table 3, top rows). For Experiment 1, there was also a significant interaction that was characterized using separate models for pre- and post-divergence intervals (Table 3, middle and bottom rows). The proportion of looking to alternate objects was significantly greater for changed activities that were subsequently recollected as such (green line) than those not recollected as such (purple line). This difference was greater in the pre- than post-divergence interval. For Experiment 2, the additional control condition prompted pairwise comparisons to characterize the effect of Change Recollection. Viewers looked to alternate objects more during activities for which changes were subsequently recollected (green line) than during activities for which changes were not recollected (purple line), z ratio = 6.84, $p < .001$, and during control activities (black line), z ratio = 7.45, $p < .001$. This pattern was highly consistent across activities varying in pre-divergence interval length (see Supplemental Material). These results suggest that mnemonic predictive looking errors play a critical role in the facilitation in memory updating that occurred when changes were recollected.

For Review Only

Proportions of looking to target objects in the Day 2 movie (Figure 5, right panels) were also compared. These analyses tested the hypothesis that viewers should not show greater predictive looking associated with subsequent change recollection because target objects had not been contacted in the Day 1 movie. Thus, mnemonic predictions should not guide the eyes to look earlier and more often to objects that the actor had not manipulated. Models comparable to those for alternate objects confirmed this prediction (Table 4). In Experiment 1, neither the effect of Change Recollection, nor Change Recollection × Bin interaction was significant. In Experiment 2, there was a significant effect of Change Recollection, but pairwise comparisons indicated no significant differences among conditions, smallest z ratio = 2.31, p = .054. Overall, these results are consistent with the hypothesis that memory only predictively guides the eyes to previous action targets. Finally, we also found that overall correct recall of changed activity features from the Day 2 movie was associated with greater proportions of looking to alternate objects in the pre-divergence interval (see Supplemental Material).

Discussion

Two experiments examined the role of mnemonic prediction error in event memory updating by assessing the association between predictive looking and memory for changed activities. Predictive looking to objects an actor had contacted was associated with facilitated memory for changed activity features. This was shown by an association between memory for earlier-contacted objects and looking to those objects during repeated actions as well as better memory for changed objects and recollection of the change. These results suggest that mnemonic prediction errors partly contribute to event memory updating for dynamic everyday actions.

The present results converge with action observation studies showing that viewers look ahead when performing and watching everyday actions (Eisenberg et al., 2018; Hayhoe et al.,

For Review Only

2003; Land & McLeod, 2000). These results replicated findings of predictive looking during movies of everyday activities (Eisenberg et al., 2018), as viewers looked increasingly more to target objects when watching Day 1 movies. These findings support the perspective that viewers use motor knowledge to predict action goals (Flanagan & Johansson, 2003). The present results also converged with findings from simplified trial-based designs showing that viewers look ahead to earlier-watched actions (Cannon et al., 2012; Falck-Ytter et al., 2006; Gredebäck et al., 2018). Specifically, when watching the Day 2 movie, viewers looked more to objects that had been contacted in the Day 1 movie. This was true for activities that repeated completely and those included only repetitions of the initial action sequence in the pre-divergence interval, suggesting that memory guided action predictions during activity viewing.

The present study also showed that mnemonic prediction errors enhanced adults' new learning. These results extend on action observation studies with infants and children showing that prediction errors invite active learning (Gredebäck et al., 2018; Juvrud et al., 2019; Stahl & Feigenson, 2015). Viewers in the preset study looked earlier and more often to repeated pre-divergence actions and changed post-divergence actions, suggesting that mnemonic prediction errors stimulated encoding of changed features. This view was also supported by the finding that change recollection, which was strongly associated with facilitated memory updating, was also associated with predictive looking to earlier-contacted objects. The present results therefore show that mnemonic prediction errors facilitate encoding and retrieval of everyday events.

The role of prediction error in event memory updating has been studied in various ways, leading to different outcomes and theoretical perspectives. Some theories assume that prediction errors impair memory when unfulfilled expectations lead to weakened representations of expected events (Kim et al., 2014) or increases in intrusion errors (Sinclair & Barense, 2018).

For Review Only

These views are similar to models proposing that reactivating existing memories prior to new events destabilizes those memories (Exton-McGuinness et al., 2015; Lee et al., 2017). Other theories propose that repeating event features triggers retrieval of existing memories, leading to predictions that associated stimuli will follow. However, when changes occur, prediction errors stimulate encoding (Antony et al., 2020; Bein et al., 2020; Chen et al., 2015; Greve et al., 2017; Wahlheim & Zacks, 2019). The present findings are consistent with the latter perspective, with evidence from looking proportions associated with error-driven updating converging with results showing that neural reactivation of dynamic naturalistic events is associated with improved encoding of changed activity features (Stawarczyk et al., 2020).

One potential mechanism for the error-driven updating observed here is integrative encoding of event representations. Interference theories predict that event changes should impair memory (Anderson & Neely, 1996). However, there is mounting evidence that interference is mitigated when prior events are retrieved and integrated with new events (Chanales et al., 2019; Wahlheim & Jacoby, 2013). This integration may be mediated partly by medial temporal structures that are biased towards encode states by mnemonic prediction errors (Bein et al., 2020). Another mechanism that could operate simultaneously is integrative encoding of eye movement shifts from first- to second-contacted objects after prediction errors. This is consistent with the view that eye movements can become embedded in and support subsequent memory (Ryan et al., 2020; Wynn et al., 2019). Such gaze-enhanced encoding could be examined by restricting eye movements (e.g., Henderson et al., 2005) to determine if preventing shifts between objects impairs memory for recently contacted objects. Finally, although we have interpreted the present looking patterns as showing that updating benefited from retrieval of activities before changes appeared, it may have also benefitted from retrospective comparisons

For Review Only

of activity features occurring after changes appeared. Research is needed to understand how the timing of retrieval-and-comparison processes affects subsequent memory updating.

Although the current paradigm is more naturalistic than many other memory updating paradigms, some features of the task limit its generalizability. Participants watched the actor perform everyday actions in a distraction-limited environment, which differs from everyday life in which viewing goals are dynamic and susceptible to distractions. Also, the present movies included an unfamiliar actor with specific demographic characteristics, which differs from everyday life because viewers also observe actions of actors of varying familiarity from various backgrounds. Future studies should manipulate these variables to better assess the conditions under which mnemonic prediction errors are associated with facilitation in memory updating.

In summary, while watching an actor perform naturalistic everyday actions, viewers looked ahead to earlier-contacted objects before action changes were perceived. This predictive looking occurred more for actions subsequently recollected to have changed, and this was associated with better memory updating for changed actions. These results support the view that mnemonic prediction errors play a critical role in event memory updating by promoting more effective encoding and retrieval of changed activity features.

For Review Only

Action Editor

Karen Rodrigue

Editor

Patricia J. Bauer

Author Contributions

C.N.W and M.L.E. conceptualized the study and contributed equally to the present research. J.M.Z. provided substantial guidance about the execution of both experiments. C.N.W. and J.M.Z. developed the stimuli used in both experiments. M.L.E. and C.N.W. prepared the stimuli, designed the program, and coordinated data collection for Experiment 1. D.S. prepared the stimuli, designed the program, and coordinated data collection for Experiment 2. J.M.Z, M.L.E, and C.N.W. cleaned and analyzed the data for both experiments. C.N.W. prepared the initial draft of the manuscript, and all authors provided critical comments and revisions.

Open Practices Statement

Neither of the experiments reported in this article was formally preregistered. De-identified data supporting the findings of this study, the code written to analyze these data, and the experimental stimuli are available in the Open Science Framework repository: https://osf.io/w8skp/ with the identifier doi:10.17605/OSF.IO/W8SKP.

Author Note

This project was funded by National Institute on Aging grants T32AG000030 and R21AG05231401, a National Science Foundation Graduate Research Fellowship Grant No. DGE-1143954, the Mr. and Mrs. Spencer T. Olin Fellowship, and the European Union's Horizon 2020 Research and Innovation Programme under Marie Skłodowska-Curie Grant Agreement No. 798109. For their assistance with data collection and coding, we thank Shelly Gupta, Katie

For Review Only

Munson, Connor Perkins, and Keshav Jayaraman. For her assistance with experiment development and data collection, we thank Hannah Aaron. Portions of the data reported here were presented at the 57th, 60th, and 61st annual meetings of the Psychonomic Society and the 2017 North Carolina Cognition Conference.

For Review Only

References

Anderson, M. C., & Neely, J. H. (1996). Interference and inhibition in memory retrieval. In E. L. Bjork & R. A. Bjork (Eds.), *Memory* (pp. 237–313). San Diego, CA: Academic Press.

Antony, J. W., Hartshorne, T. H., Pomeroy, K., Gureckis, T. M., Hasson, U., McDougle, S. D., & Norman, K. A. (2020). Behavioral, physiological, and neural signatures of surprise during naturalistic sports viewing. *Neuron*. https://doi.org/10.1016/j.neuron.2020.10.029

Barnes, G. R., Collins, C. J. S., & Arnold, L. R. (2005). Predicting the duration of ocular pursuit in humans. *Experimental Brain Research*, *160*(1), 10–21. https://doi.org/10.1007/s00221-004-1981-3

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using **lme4**. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bein, O., Duncan, K., & Davachi, L. (2020). Mnemonic prediction errors bias hippocampal states. *Nature Communications*, *11*(1), 3451. https://doi.org/10.1038/s41467-020-17287-1

Cannon, E. N., Woodward, A. L., Gredebäck, G., Hofsten, C. von, & Turek, C. (2012). Action production influences 12-month-old infants' attention to others' actions. *Developmental Science*, *15*(1), 35–42. https://doi.org/10.1111/j.1467-7687.2011.01095.x

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), 5363. https://doi.org/10.1038/s41467-019-13377-x

Chen, J., Cook, P. A., & Wagner, A. D. (2015). Prediction strength modulates responses in human area CA1 to sequence violations. *Journal of Neurophysiology*, *114*(2), 1227–1238. https://doi.org/10.1152/jn.00149.2015

For Review Only

Diaz, G., Cooper, J., Rothkopf, C., & Hayhoe, M. (2013). Saccades to future ball location reveal memory-based prediction in a virtual-reality interception task. *Journal of Vision*, *13*(1), 1–14. https://doi.org/10.1167/13.1.20

Eisenberg, M. L., Zacks, J. M., & Flores, S. (2018). Dynamic prediction during perception of everyday events. *Cognitive Research: Principles and Implications*, *3*(1), 53. https://doi.org/10.1186/s41235-018-0146-z

Exton-McGuinness, M. T. J., Lee, J. L. C., & Reichelt, A. C. (2015). Updating memories—The role of prediction errors in memory reconsolidation. *Behavioural Brain Research*, *278*, 375–384. https://doi.org/10.1016/j.bbr.2014.10.011

Falck-Ytter, T., Gredebäck, G., & von Hofsten, C. (2006). Infants predict other people's action goals. *Nature Neuroscience*, *9*(7), 878–879. https://doi.org/10.1038/nn1729

Flanagan, J. R., & Johansson, R. S. (2003). Action plans used in action observation. *Nature*, *424*(6950), 769–771. https://doi.org/10.1038/nature01861

Fox, J., & Weisberg, S. (2019). *An {R} Companion to Applied Regression, Third Edition.* Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Garlitch, S. M., & Wahlheim, C. N. (2021). Directing attention to event changes improves memory updating for older adults. *Psychology and Aging*, *36*(4), 475–490. https://doi.org/10.1037/pag0000503

Gerson, S. A., & Woodward, A. L. (2014). Learning from their own actions: The unique effect of producing actions on infants' action understanding. *Child Development*, *85*(1), 264–277. https://doi.org/10.1111/cdev.12115

For Review Only

Gredebäck, G. (2018). How visual and motor experience shapes the development of infants' perception of actions performed by social partners. *Journal of Motor Learning and Development*, *6*(s1), S89–S104. https://doi.org/10.1123/jmld.2016-0074

Gredebäck, G., & Falck-Ytter, T. (2015). Eye movements during action observation. *Perspectives on Psychological Science*, *10*(5), 591–598. https://doi.org/10.1177/1745691615589103

Gredebäck, G., Lindskog, M., Juvrud, J. C., Green, D., & Marciszko, C. (2018). Action prediction allows hypothesis testing via internal forward models at 6 months of age. *Frontiers in Psychology*, *9*, 290. https://doi.org/10.3389/fpsyg.2018.00290

Greve, A., Cooper, E., Kaula, A., Anderson, M. C., & Henson, R. (2017). Does prediction error drive one-shot declarative learning? *Journal of Memory and Language*, *94*, 149–165. https://doi.org/10.1016/j.jml.2016.11.001

Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision*, *3*(1), 6. https://doi.org/10.1167/3.1.6

Henderson, J. M., Williams, C. C., & Falk, R. J. (2005). Eye movements are functional during face learning. *Memory & Cognition*, *33*(1), 98–106.

Hermann, M. M., Wahlheim, C. N., Alexander, T. R., & Zacks, J. M. (2021). The role of prior-event retrieval in encoding changed event features. *Memory & Cognition*. https://doi.org/10.3758/s13421-021-01173-2

Juvrud, J., Bakker, M., Kaduk, K., DeValk, J. M., Gredebäck, G., & Kenward, B. (2019). Longitudinal continuity in understanding and production of giving-related behavior from infancy to childhood. *Child Development*, *90*(2), e182–e191. https://doi.org/10.1111/cdev.13131

For Review Only

Kafkas, A., & Montaldi, D. (2018). Expectation affects learning and modulates memory experience at retrieval. *Cognition*, *180*, 123–134. https://doi.org/10.1016/j.cognition.2018.07.010

Kim, G., Lewis-Peacock, J. A., Norman, K. A., & Turk-Browne, N. B. (2014). Pruning of memories by context-based prediction error. *Proceedings of the National Academy of Sciences*, *111*(24), 8997–9002. https://doi.org/10.1073/pnas.1319438111

Land, M. F., & McLeod, P. (2000). From eye movements to actions: How batsmen hit the ball. *Nature Neuroscience*, *3*(12), 1340–1345. https://doi.org/10.1038/81887

Lee, J. L. C., Nader, K., & Schiller, D. (2017). An update on memory reconsolidation updating. *Trends in Cognitive Sciences*, *21*(7), 531–545. https://doi.org/10.1016/j.tics.2017.04.006

Lenth, R. (2020). *emmeans: Estimated Marginal Means, aka Least-Squares Means.* (R package version 1.4.7.) [Computer software]. https://CRAN.R-project.org/package=emmeans

Melzer, A., Prinz, W., & Daum, M. M. (2012). Production and perception of contralateral reaching: A close link by 12 months of age. *Infant Behavior and Development*, *35*(3), 570–579. https://doi.org/10.1016/j.infbeh.2012.05.003

Neely, J. H., Keefe, D. E., & Ross, K. L. (1989). Semantic priming in the lexical decision task: Roles of prospective prime-generated expectancies and retrospective semantic matching. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(6), 1003–1019. https://doi.org/10.1037/0278-7393.15.6.1003

R Core Team. (2020). *R Software*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement learning and nonreinforcement. In A. H. Black & W. F.

For Review Only

Prokasy (Eds.), *Classical Conditioning II* (pp. 64–88). Appleton-Century-Crofts: New York.

Ryan, J. D., Shen, K., & Liu, Z. (2020). The intersection between the oculomotor and hippocampal memory systems: Empirical developments and clinical implications. *Annals of the New York Academy of Sciences*, *1464*(1), 115–141. https://doi.org/10.1111/nyas.14256

Sinclair, A. H., & Barense, M. D. (2018). Surprise and destabilize: Prediction error influences episodic memory reconsolidation. *Learning & Memory*, *25*(8), 369–381. https://doi.org/10.1101/lm.046912.117

Stahl, A. E., & Feigenson, L. (2015). Observing the unexpected enhances infants' learning and exploration. *Science*, *348*(6230), 91–94. https://doi.org/10.1126/science.aaa3799

Stawarczyk, D., Wahlheim, C. N., Etzel, J. A., Snyder, A. Z., & 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*, *117*(47), 29346–29353.

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

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

Wynn, J. S., Shen, K., & Ryan, J. D. (2019). Eye movements actively reinstate spatiotemporal mnemonic content. *Vision*, *3*(2), 21. https://doi.org/10.3390/vision3020021

For Review Only

Table 1

Looking Model Results for Each Activity Type: Experiments 1 and 2

		Experiment 1			Experiment 2		
Activity	Effect	χ^2	*df*	*p*	χ^2	*df*	*p*
Repeated	Object	5216.37	1	< .001	5956.10	1	< .001
	Day	31.31	1	< .001	23.44	1	< .001
	Bin	1387.58	19	< .001	1406.82	19	< .001
	Object × Day	73.41	1	< .001	73.88	1	< .001
	Day × Bin	7.76	19	= .989	7.53	19	= .991
	Object × Bin	4885.20	19	< .001	5353.73	19	< .001
	Object × Day × Bin	17.88	19	= .530	42.82	19	= .001
Changed	Object	4760.08	1	< .001	9733.85	1	< .001
	Day	59.77	1	< .001	33.70	1	< .001
	Bin	1336.73	19	< .001	3092.08	19	< .001
	Object × Day	24.99	1	< .001	0.75	1	= .386
	Day × Bin	10.24	19	= .947	3.26	19	= .999
	Object × Bin	5100.31	19	< .001	11544.90	19	< .001
	Object × Day × Bin	79.34	19	< .001	28.79	19	= .069

Note. The model results correspond to the data displayed in Figure 3.

For Review Only

Table 2

Looking Model Results for Repeated Activities in Experiment 2 and Changed Activities in Experiment 1

Activity	Experiment	Interval	Effect	χ^2	*df*	*p*
Repeated	Experiment 2	Pre-Divergence	Object	19.93	1	< .001
			Day	17.74	1	< .001
			Bin	55.97	9	< .001
			Object × Day	6.45	1	= .011
			Day × Bin	7.54	9	= .581
			Object × Bin	65.84	9	< .001
			Object × Day × Bin	8.81	9	= .455
		Post-Divergence	Object	9415.61	1	< .001
			Day	8.93	1	= .003
			Bin	89.84	9	< .001
			Object × Day	83.43	1	< .001
			Day × Bin	1.27	9	= .999
			Object × Bin	401.91	9	< .001
			Object × Day × Bin	7.67	9	= .568
Changed	Experiment 1	Pre-Divergence	Object	0.20	1	= .651
			Day	31.78	1	< .001
			Bin	43.95	9	< .001
			Object × Day	8.09	1	= .004
			Day × Bin	6.43	9	= .697
			Object × Bin	35.33	9	< .001
			Object × Day × Bin	10.85	9	= .286
		Post-Divergence	Object	8035.43	1	< .001
			Day	29.01	1	< .001
			Bin	117.62	9	< .001
			Object × Day	76.63	1	< .001
			Day × Bin	3.35	9	= .949
			Object × Bin	517.59	9	< .001
			Object × Day × Bin	0.92	9	= .999

Note. The model results correspond to the data displayed in Figure 3.

For Review Only

Table 3

Model Results for Day 2 Looking to Alternate Objects Conditionalized on Change Recollection

Experiment	Time Bins	Effect	χ^2	*df*	*p*
Experiment 1	All	Change Recollection	49.15	1	< .001
		Bin	652.95	19	< .001
		Change Recollection × Bin	34.91	19	= .014
	Pre-Divergence	Change Recollection	52.27	1	< .001
		Bin	15.34	9	= .082
		Change Recollection × Bin	3.40	9	= .946
	Post-Divergence	Change Recollection	7.57	1	= .006
		Bin	73.12	9	< .001
		Change Recollection × Bin	5.27	9	= .810
Experiment 2	All	Change Recollection	60.87	2	< .001
		Bin	1611.21	19	< .001
		Change Recollection × Bin	41.22	38	= .332

Note. The model results correspond to the data displayed in Figure 5 (left panels). For Experiment 2, note that the Change Recollection variable include both changed activities conditionalized on change recollection and control activities that were not conditionalized on change recollection.

For Review Only

Table 4

Model Results for Day 2 Looking to Targets Conditionalized on Change Recollection

Experiment	Effect	χ^2	*df*	*p*
Experiment 1	Change Recollection	2.04	1	= .153
	Bin	2940.44	19	< .001
	Change Recollection × Bin	19.17	19	= .446
Experiment 2	Change Recollection	6.40	2	= .041
	Bin	9022.16	19	< .001
	Change Recollection × Bin	24.37	38	= .958

Note. The model results correspond to the data displayed in Figure 5 (right panels). For Experiment 2, note that the Change Recollection variable include both changed activities conditionalized on change recollection and control activities that were not conditionalized on change recollection.

For Review Only

Figure Captions

Fig. 1. *Example Activities and Interest Areas*

Note. A) Key moments from an example activity in which the actor unlocked a door. The interest period for eye tracking analysis began as she approached the door and ended when she first contacted either the *doorknob* (Version A) or *deadbolt* (Version B). The first images (Start) show the start of the interest period; the second images (Divergence) show the moment before the activity versions diverge; the third images (Contact) show the two possible contacted objects. B) Example interest areas for a repeated activity. C) Example interest areas for a changed activity. B and C) Yellow boxes are interest areas for target (blue) and alternate (red) objects. The designation of target and alternate objects was the same in both movies for repeated activities and switched from the Day 1 to Day 2 movie for changed activities. The images are shown in black-and-white here to enhance the visual contrast of the yellow boxes showing the interest areas. The complete movies were shown in color and were not altered to indicate interest areas. The movies can be downloaded from the OSF: https://osf.io/w8skp/.

Fig. 2. *General Schematic of Events in Each Experimental Session*

Note. A) Participants watched two movies successively during the first session. They were instructed to pay attention to the actor's activities in both movies. After a delay, participants completed (B) a second session that included a cued recall test of features from both movies. For each activity, participants first tried to recall the Day 2 feature. Then they were asked to classify if the activity had earlier changed between movies. When participants indicated that an activity had changed, they tried to recall the Day 1 feature then moved on to the next trial (blue arrows). When participants indicated that an activity had not changed, the program advanced to the next

For Review Only

test trial (red arrow). The test instructions differed slightly between experiments (see Supplementary Material).

Fig. 3. *Looks to Target and Alternate Objects During Movie Viewing*

Note. Model-estimated proportions of looking in interest areas including target and alternate objects during interest periods for critical activities. “Start” indicates the moment the interest period began. “Divergence” indicates the point of divergence when the two activity versions first became distinguishable. “Contact” indicates the moment the actor contacted the target or alternate object. Orange dashed lines indicate looking proportions during Day 1 movie viewing. Black solid lines indicate looking proportions during Day 2 movie viewing.

Fig. 4. *Cued Recall of Activity Features*

Note. Black points indicate model-estimated probabilities for overall responses. Colored points for conditional cells indicate model-estimated response probabilities for changed activities when changes were recollected (green points), remembered (blue points), or not remembered (red points). Point area differences indicate the relative observation frequencies for those cells. Overall correct recall of Day 2 activity features was better for repeated than changed and for changed than control activities (left panels). Changed activities were recalled best when change was recollected. Intrusions of alternate activity features for changed activities were greater when changes were remembered than when they were not (right panels). Error bars are 95% confidence intervals. Error bars are obscured when the intervals are smaller than the point areas.

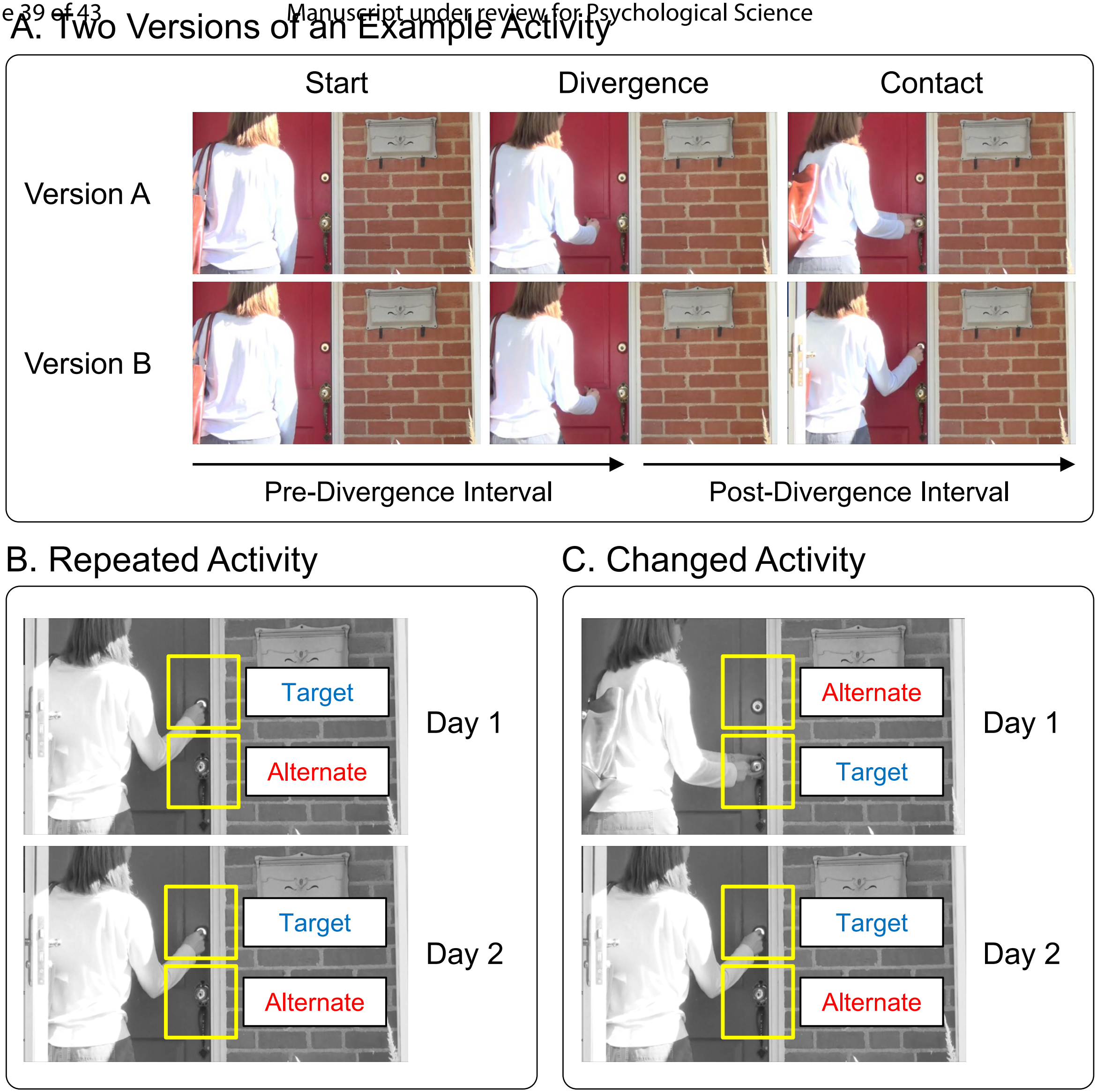
Fig. 5. *Looks to Interest Areas during Changed Activities on Day 2*

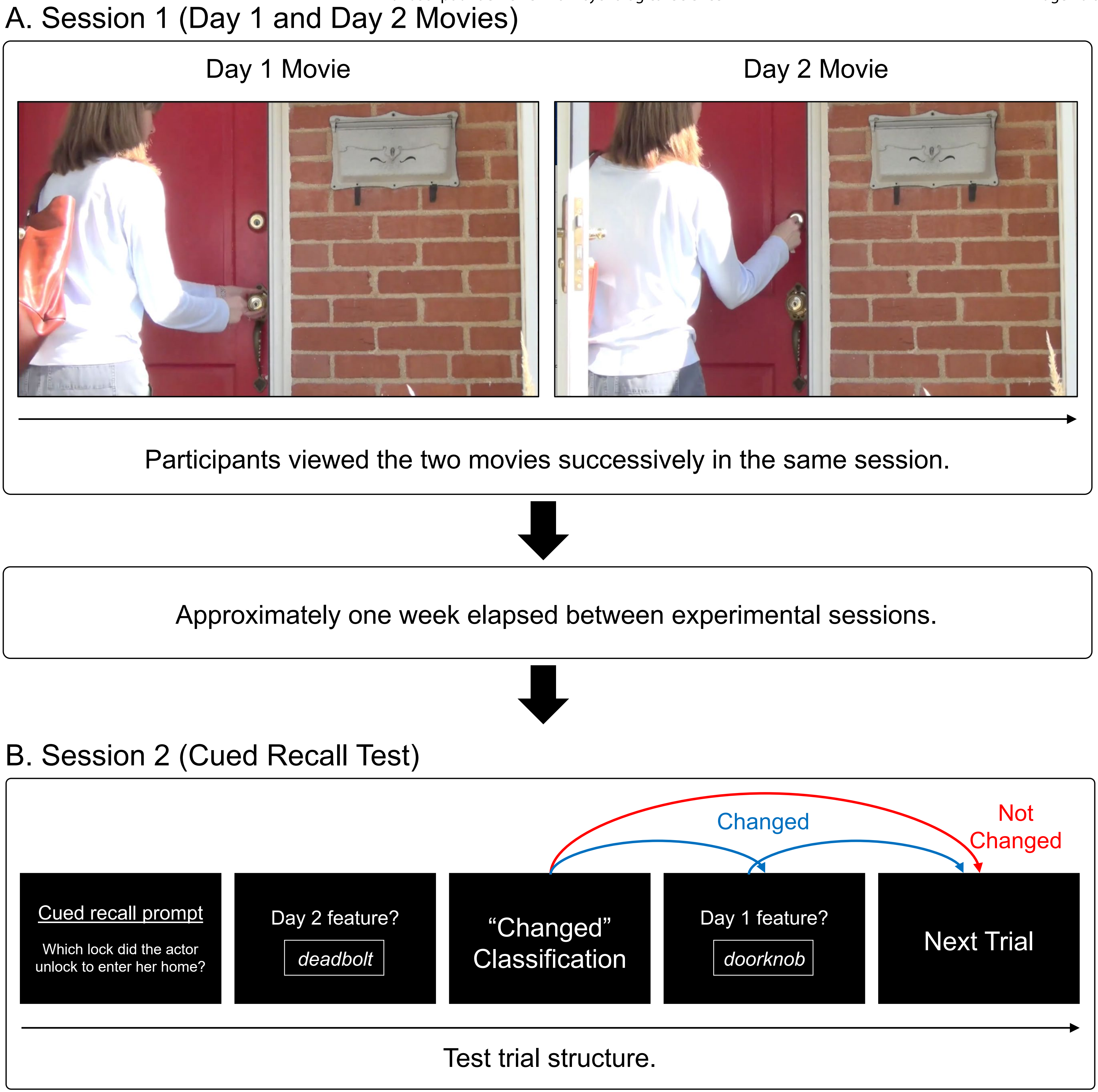
Note. Lines indicate model-estimated proportions of looking to interest areas for changed activities during Day 2 movies in Experiments 1 and 2. Green solid lines indicate looking proportions for activities recollected as changed. Purple solid lines indicate looking proportions

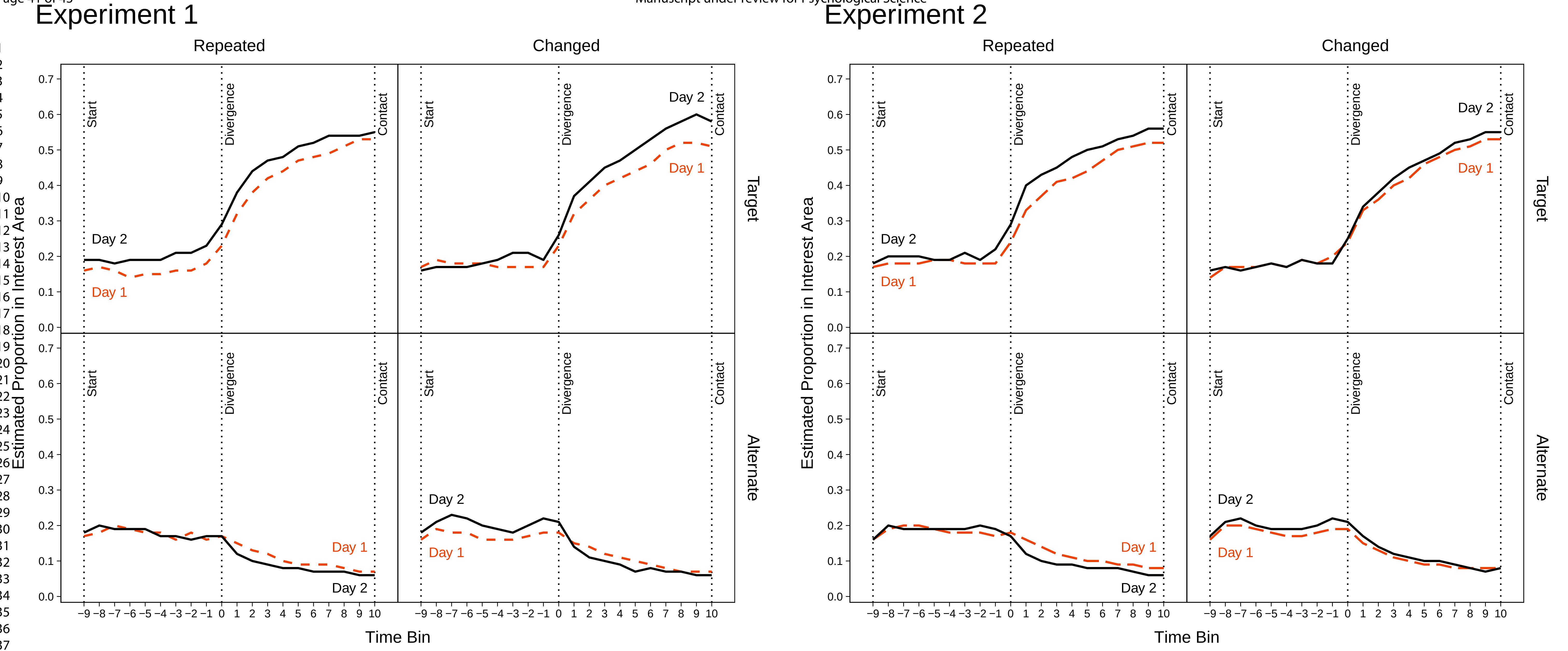
For Review Only

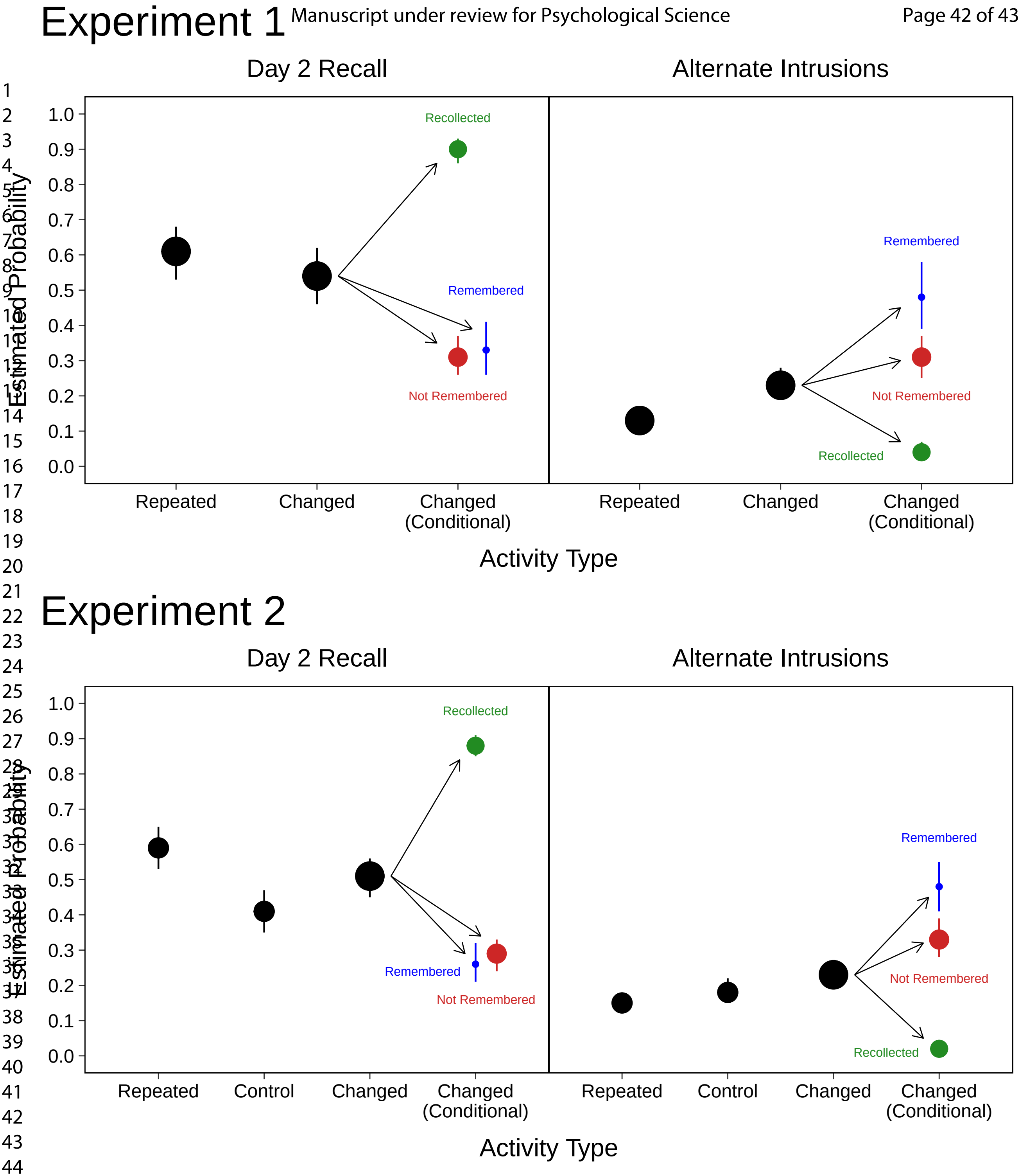
for activities not recollected as changed. Black dotted lines in Experiment 2 indicate looking proportions for control activities. Viewers looked at alternate objects more during pre-divergence intervals for changed activities subsequently recollected as such, showing that mnemonic predictive looking errors were associated with facilitation in event memory updating.

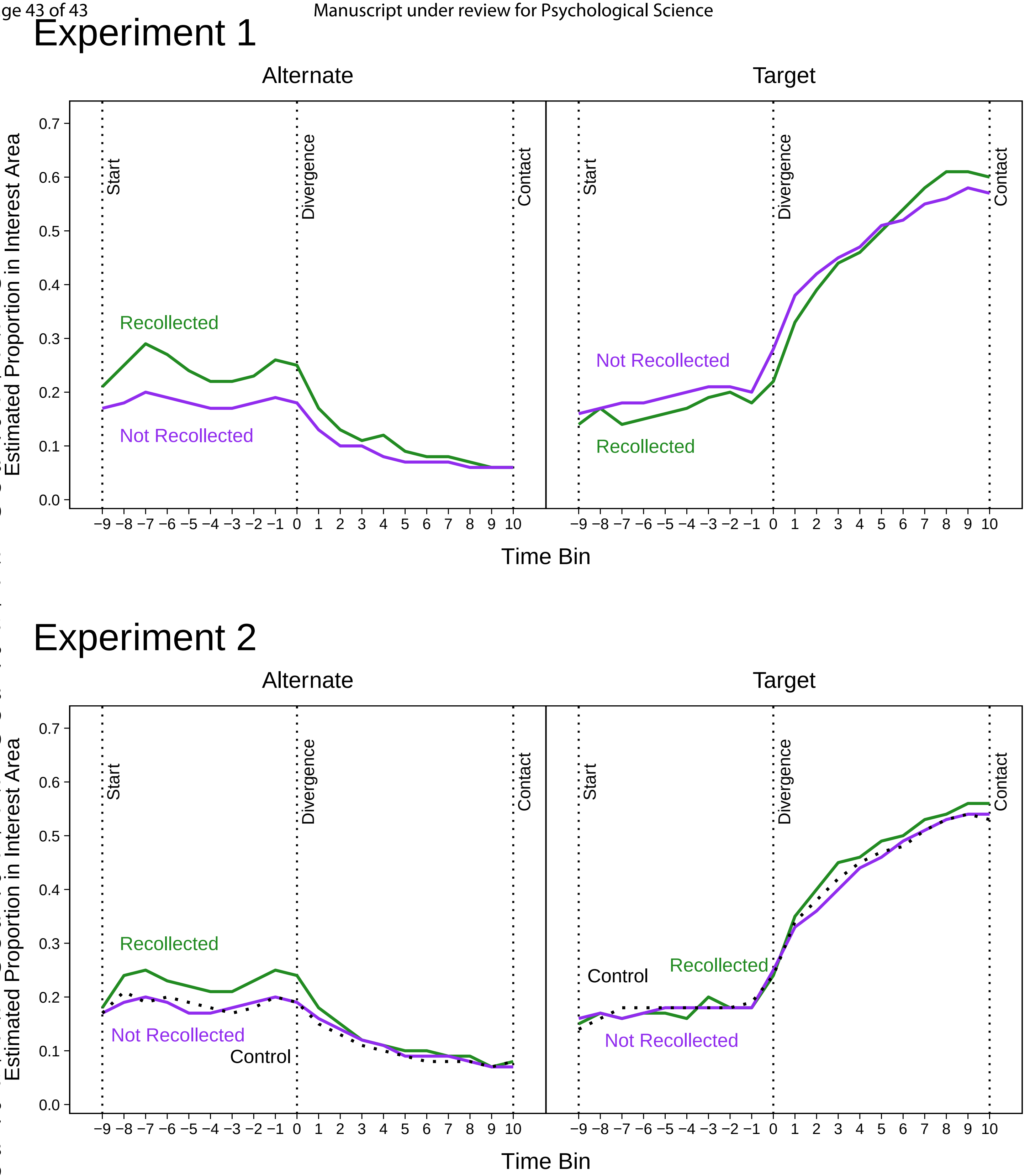
For Review Only


A. Two Versions of an Example Activity
Start
Divergence
Contact
Version A
Version B
Pre-Divergence Interval
Post-Divergence Interval
B. Repeated Activity
Target
Alternate
Day 1
Target
Alternate
Day 2
C. Changed Activity
Alternate
Target
Day 1
Target
Alternate
Day 2


A. Session 1 (Day 1 and Day 2 Movies)
Day 1 Movie
Day 2 Movie
Participants viewed the two movies successively in the same session.
Approximately one week elapsed between experimental sessions.
B. Session 2 (Cued Recall Test)
Changed
Not Changed
Cued recall prompt
Which lock did the actor unlock to enter her home?
Day 2 feature?
deadbolt
“Changed” Classification
Day 1 feature?
doorknob
Next Trial
Test trial structure.


Experiment 1
Experiment 2
Repeated
Changed
Target
Alternate
Start
Divergence
Contact
Day 1
Day 2
Estimated Proportion in Interest Area
Time Bin
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
−9 −8 −7 −6 −5 −4 −3 −2 −1 0 1 2 3 4 5 6 7 8 9 10


Experiment 1
Day 2 Recall
Alternate Intrusions
Estimated Probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Recollected
Remembered
Not Remembered
Repeated
Changed
Changed (Conditional)
Activity Type
Experiment 2
Day 2 Recall
Alternate Intrusions
Estimated Probability
Recollected
Remembered
Not Remembered
Repeated
Control
Changed
Changed (Conditional)
Activity Type


Experiment 1
Alternate
Target
Estimated Proportion in Interest Area
Start
Divergence
Contact
Recollected
Not Recollected
Not Recollected
Recollected
Time Bin
Experiment 2
Alternate
Target
Estimated Proportion in Interest Area
Start
Divergence
Contact
Recollected
Not Recollected
Control
Control
Recollected
Not Recollected
Time Bin