Research Paper

Interpolated retrieval retroactively increases recall and promotes cross-episode memory interdependence

Christopher N. Wahlheim,[1] Sydney T. Smith,[1] Sydney M. Garlitch,[2] and Robert W. Wiley[1]

[1] *Department of Psychology, University of North Carolina at Greensboro, Greensboro, North Carolina 27402, USA*
[2] *Behavioral Sciences Department, Millikin University, Decatur, Illinois 62522, USA*

Retrieving existing memories before new learning can lead to retroactive facilitation. Three experiments examined whether interpolated retrieval is associated with retroactive facilitation and memory interdependence that reflects integrative encoding. Participants studied two lists of cue–response word pairs that repeated across lists (A–B, A–B), appeared in list 1 (A–B, —), or included the same cues with changed responses in each list (A–B, A–C). For A–B, A–C pairs, the tasks interpolated between lists required recall of list 1 (B) responses (with or without feedback) or restudy of complete list 1 (A–B) pairs. In list 2, participants only studied pairs (experiment 1) or studied pairs, attempted to detect changed (C) responses, and attempted to recall list 1 responses for detected changes (experiments 2 and 3). On a final cued recall test, participants attempted to recall list 1 responses, indicated whether responses changed between lists, and if so, attempted to recall list 2 responses. Interpolated retrieval was associated with subsequent retroactive facilitation and greater memory interdependence for B and C responses. These correlational findings are compatible with the view that retrieval retroactively facilitates memories, promotes coactivation of existing memories and new learning, and enables integrative encoding that veridically binds information across episodes.

[Supplemental material is available for this article.]

People experience changes every day—colleagues change their opinions, flight times change due to inclement weather, and hiking trails close for construction. Adapting to changes can entail displacing existing memories to prioritize recent events, but sometimes people need to remember how existing memories relate to subsequent changes. For example, plans to visit a hiking trail after construction has ended can be maintained by representing both the initial desire to visit the trail and the construction end date. However, conflicting details of new episodes can retroactively compete with existing memories (Müller and Pilzecker 1900). Such competition can be mitigated by differentiating contexts associated with conflicting episodes (for reviews, see Abra 1972; Smith and Vela 2001), which may occur when retrieving existing memories before a new episode shifts mental context (for reviews, see Pastötter and Bäuml 2014; Chan et al. 2018a; Yang et al. 2018). Existing memories can also be enhanced when new episodes with shared features cue awareness of the relationship (Bruce and Weaver 1973; Robbins and Bray 1974), increase the accessibility of existing memories, and enable encoding of cross-episode associations (for review, see Wahlheim et al. 2021). Accordingly, retrieving existing memories before new related episodes may improve existing memories by enhancing awareness of the relationship between episodes.

The present study examined the role of interpolated retrieval in the retroactive enhancement of existing memories and subsequent cross-episode integrative encoding using A–B, A–C tasks (for review, see Anderson and Neely 1996). This approach is ideal because the stimuli include shared and unique features that allow for assessment of accessibility and interdependence in responses across phases (Wahlheim and Jacoby 2013; Jacoby et al. 2015; Negley et al. 2018; Garlitch and Wahlheim 2020). In these tasks, participants study two lists of cue–response pairs with varying relationships among elements between lists. The conditions can include the same pairs in both lists (A–B, A–B), pairs that only appear in list 1 as controls (A–B), and pairs with the same cue and changed responses across lists (A–B, A–C). With this design, interpolated tasks between lists can be manipulated, and subsequent effects on cued recall of list 1 (B) and then list 2 (C) responses can be assessed. Retroactive memory effects are shown when list 1 recall for A–B, A–C pairs is higher than for control pairs (facilitation) or lower than for control pairs (interference). Many variables determine how new learning episodes will recall list 1 responses. One key feature is whether task characteristics prevent awareness of relationships across lists, thereby differentiating contexts, or promote awareness of relationships, thus enabling the integration of new episodes with existing memories. We use the term integration here to describe an encoding mechanism that veridically associates information from different sources (i.e., cross-episode associations). Integration can also be used to describe memory distortions that occur via overgeneralization (Warren et al. 2014) or misattributions (Gershman et al. 2013), but we restricted our conceptualization to responses that are accurately represented together.

One view of retroactive memory effects is that the encoding and retrieval of contextual associations determine the extent to which exposure to conflicting events leads to interference (e.g., Mensink and Raaijmakers 1988). Accordingly, people store information about items as well as their associated mental and physical contexts (Bower 1972; Martin 1972). When retrieving existing memories, the contexts of new learning and existing memories are not well distinguished, leading to response competition that

Corresponding author: cnwahlhe@uncg.edu
Article is online at http://www.learnmem.org/cgi/doi/10.1101/lm.053782.123.

reduces memory accuracy. Therefore, conditions that promote the differentiation of contexts associated with competing information should prevent retroactive interference. Indeed, many studies have accomplished interference reduction by increasing differences in the temporal and physical contexts associated with related episodes (for reviews, see Abra 1972; Smith and Vela 2001). In addition, neural studies of interference reduction suggest that prefrontal (e.g., Henson et al. 2002) and hippocampal (e.g., Favila et al. 2016) regions support the encoding of new information as distinct from existing memories. It follows that perfect differentiation should lead to no interdependence between episodic memory representations, thus eliminating interference but not producing facilitation.

In contrast, the integrative encoding view proposes that facilitation occurs when perceptual inputs trigger study-phase retrievals of existing memories, leading to coactivation that supports the formation of cross-episode associations (e.g., Hintzman 2011; Schlichting and Preston 2015). Accordingly, integrative encoding serves to preserve the relative differences in associated contexts. This not only prevents interference by allowing participants to remember which episode was the reminder, and therefore more recent, but also leads to retroactive facilitation driven by retrieval practice and elaborated re-encoding of existing memories (Jacoby et al. 2015). Evidence that stimulus similarity can enhance existing memories was shown when paired associates with similar responses (e.g., afraid and scared) produced retroactive facilitation (Barnes and Underwood 1959). Retroactive facilitation was also shown with less similar responses when participants were aware of relationships across lists (Bruce and Weaver 1973; Robbins and Bray 1974), which may have induced study-phase retrievals of list 1 responses during list 2 study (Benjamin and Ross 2010). In this way, integrative encoding enabled by retrieval of existing memories can support the differentiation of context associated with elements in cross-episode associations. This idea suggests that although context differentiation and integrative encoding accounts have been considered separately, the dichotomy may be false under some circumstances (cf. Kumaran and McClelland 2012).

The role of study-phase retrievals in retroactive facilitation was confirmed using a looking back procedure in an A–B, A–C task that varied study-phase retrievals during list 2 (Jacoby et al. 2015; see also Jacoby 1974; Jacoby and Wahlheim 2013). Instructing participants to indicate when responses in list 2 had changed from list 1 led to retroactive facilitation in subsequent list 1 recall and better memory that responses had changed (an indirect measure of memory interdependence). Related evidence for increased memory interdependence has been shown using other manipulations that increased study-phase retrievals. For example, repeating background contexts associated with A–B pairs during A–C encoding has been shown to produce retroactive facilitation (Cox et al. 2021) and increased interdependence relative to when different background contexts appeared with A–B and A–C pairs. Also, increasing the strength of semantic associations between B and C responses has been shown to reduce interference in list 1 recall and increase interdependence (e.g., see Osgood 1949; Antony et al. 2022). More generally, these findings are compatible with work showing that cross-episode connections enhance holistic recollection across the life span (Horner and Burgess 2014; Horner et al. 2015; Ngo et al. 2019, 2021).

Neural reactivation effects on existing memories in A–B, A–C tasks also suggest that study-phase retrievals retroactively enhance memory. For example, correlated activation in the hippocampus and frontostriatal regions during encoding of list 2 pairs was associated with better memory for list 1 pairs in cued recall, suggesting that reactivating pairs and their contexts protected existing memories against interference (Kuhl et al. 2010). Additionally, neural pattern similarity between reactivated list 1 pairs and studied list 2 pairs in regions associated with list 1 encoding tasks was associated with better memory for the mental context (encoding task) associated with list 1 pairs (Koen and Rugg 2016). Similarly, neural reactivation of encoding context measured by pattern similarity in oscillatory brain activity was associated with successful list 1 retrieval under conditions of retroactive interference (Bramão et al. 2022). These studies suggest that increasing A–B accessibility during A–C study can enhance memory for A–B pairs through retrieval practice and by enabling cross-episode associations. This is compatible with evidence for integrative encoding from related paired-associate learning tasks that we summarize in the "Discussion" section (e.g., see Schlichting and Preston 2014; Chanales et al. 2019).

Both views above have been invoked to account for interference reduction accomplished by interpolated testing. Studies of interpolated testing and interference have often used either multilist free recall or A–B, A–C tasks. An early study of interpolated retrieval effects on interference used multilist free recall (Darley and Murdock 1971). Participants studied word lists and completed recall tests after half of the lists. Prior-list intrusions were mostly from nontested lists, suggesting that interpolated retrieval mitigated interference. In a similar study, final list recall showed higher correct recall and fewer prior-list intrusions when earlier study lists had been tested as compared with restudied or not tested at all (Szpunar et al. 2008). These findings were attributed to retrieval practice increasing the contextual distinctiveness of studied items, thus supporting postretrieval monitoring. This view has been invoked to account for other retrieval types that promote differentiation (Howard and Kahana 2002; Jang and Huber 2008; Sahakyan and Hendricks 2012) and to include a role for enhanced attention (Pastötter et al. 2011). Such interference reduction cannot be accounted for by an integrative encoding view because recall tasks often use unrelated items that do not enable memory interdependence (but see Chan et al. 2018b).

Another early study of interpolated retrieval and interference used an A–B, A–C task with a procedure suited for assessing response interdependence (Tulving and Watkins 1974). Participants studied two lists of word pairs and were tested immediately after list 1, list 2, or both lists before attempting to recall responses from both lists. Interpolated testing, manipulated between and within subjects, improved subsequent list 2 recall. This was interpreted as showing that list 1 retrieval insulated it from list 2, consistent with context differentiation. However, memory interdependence was shown on the final test in that list 2 recall was greater when list 1 responses were also recalled than when they were not, indicating that the lists were not completely differentiated. A later study using a similar procedure showed increased recall of both responses when participants were told before the study that they would recall both responses and were tested on list 1 before list 2 (Arkes and Lyons 1979). For participants given interpolated retrieval, instructing them to recall list 1 during list 2 led to better final list 1 recall and more interdependence, suggesting that the list contexts were differentiated as a consequence of integrative encoding.

More recent work has also examined interpolated testing of specific list 1 pairs and proactive effects of memory using A–B, A–C tasks. In one study, participants restudied and recalled subsets of list 1 responses before studying list 2, then attempted cued recall of list 2 responses and indicated whether they recollected responses changing between lists (Wahlheim 2015). The measure of change recollection indexed differences in the accessibility of cross-episode associations because participants based these judgments partly on list 1 recall (Wahlheim et al. 2019). Interpolated recall led to higher list 2 recall and change recollection; list 2 recall was also higher when change was recollected, showing evidence for memory interdependence. A recent study replicated these findings using a final test procedure that assessed list 1 recall after participants reported recollecting changes (Kemp et al. 2023). This showed more direct evidence of memory interdependence between B and C responses. Both studies showed that successful interpolated retrieval was associated with better list 2 recall, suggesting that retrieval practice of list 1 pairs promoted study-phase retrieval of those specific pairs during list 2 study.

In the present study, we further examined associations among interpolated retrieval, memory interdependence, and recall accuracy in A–B, A–C tasks. The present study was unique in two important ways. First, recent work that assessed memory interdependence by measuring memory for changes examined the proactive effects of memory by testing list 2 responses before list 1 responses, leading to output interference on list 1 response accessibility. Here, we examined the retroactive effects of exposure to C responses in list 2 on the downstream accessibility of existing memories from list 1 (B responses) by cuing list 1 recall before list 2 recall. Second, study-phase retrievals in list 2 have been inferred from memory interdependence at test, whereas here we included overt measures of detecting changed list 2 responses (C) and list 1 response recall (B). The present tasks are therefore better suited than prior approaches for evaluating the consequences of interpolated retrieval for the accessibility of existing memories and their interdependence with new information, including shared and unique features.

Figure 1 presents an overview of the conditions (Fig. 1A) and task procedures (Fig. 1B) in the present three experiments. The A–B, A–C tasks included within-subject manipulations of interpolated test and restudy trials. After list 1 study, interpolated testing entailed presenting list 1 cues (A) and instructing participants to recall list 1 (B) responses. After list 2 study, on final test trials,

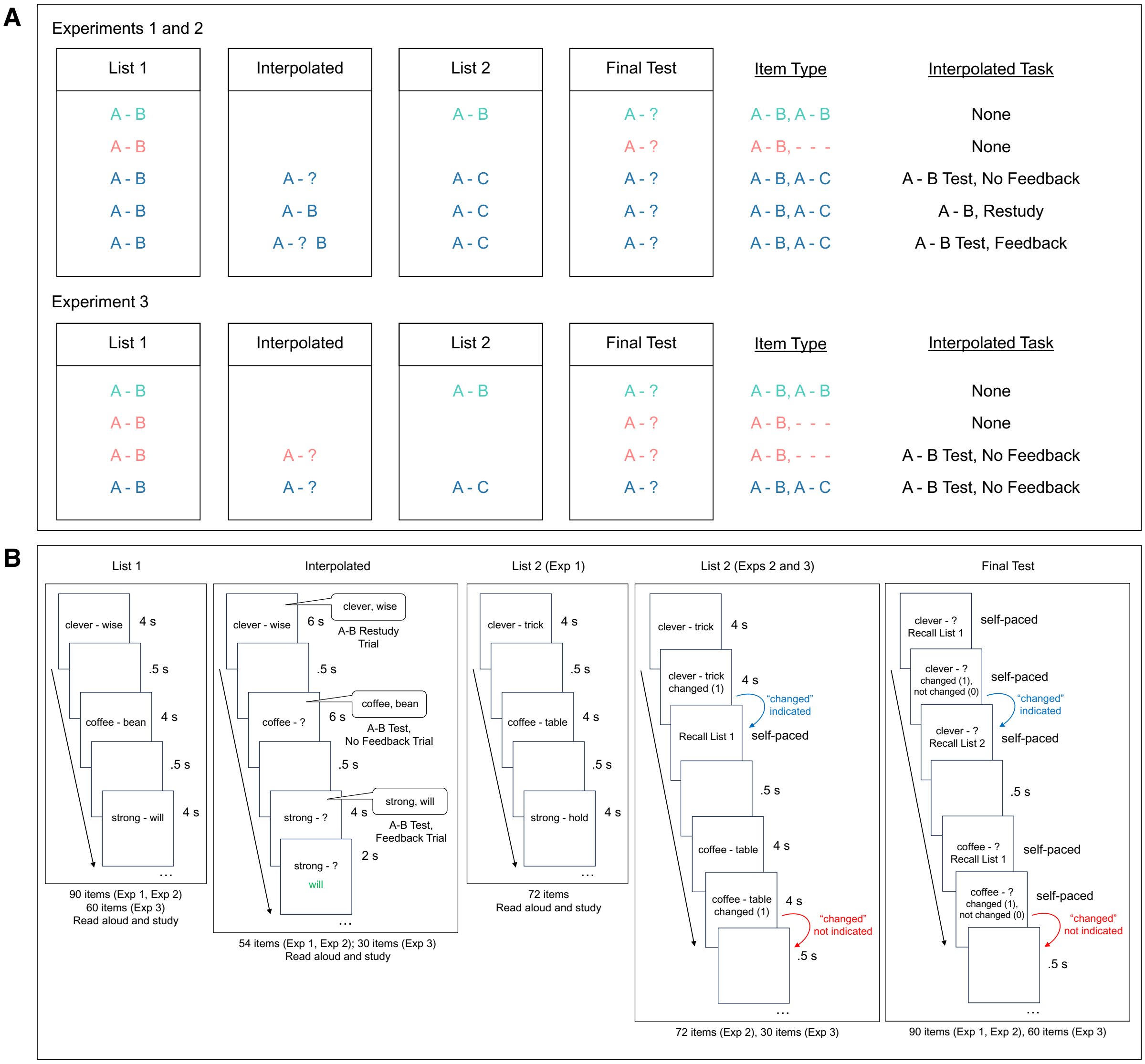

Figure 1. Schematics illustrating manipulations across conditions (*A*) and trial structures across all phases using examples of A–B, A–C item types (*B*). (*A*) Experiments 1 and 2 included five conditions, and experiment 3 included four conditions comprising combinations of varying relationships between cue–response word pairs in each list and tasks interpolated between lists. (*B*) During list 1, participants read aloud and studied word pairs. During the interpolated phase, participants restudied A–B pairs and attempted to retrieve B responses from other A–B pairs with or without feedback. During list 2, participants read aloud and studied word pairs (experiment 1) or studied word pairs, indicated whether responses changed from list 1, and if so, attempted to recall list 1 responses (experiments 2 and 3). On the final test, participants attempted to recall list 1 responses, indicated whether responses changed between lists, and if so, attempted to recall list 2 responses.

participants attempted to recall list 1 responses (B), indicated whether responses had changed, and if so, attempted to recall list 2 responses (C). We characterized retroactive memory effects on final test recall without measuring study-phase retrievals in list 2 to avoid reactive effects on A–C encoding in experiment 1. We added measures of changed response detection and list 1 recall during list 2 in experiment 2. Interpolated restudy and testing with feedback conditions in those experiments allowed us to characterize the effects of re-exposure and corrective feedback after unsuccessful retrieval. Finally, we examined interpolated retrieval effects with and without A–C study items in list 2 to determine whether response competition dampened interpolated retrieval benefits on final list 1 recall in experiment 3. This allowed us to rule out an account that interpolated testing had its effects only by promoting retrieval practice instead of also promoting study-phase retrievals that enabled cross-episode binding. In all experiments, we examined interpolated testing effects on the interdependence in recall of list 1 and 2 responses. We hypothesized that the retrieval practice of list 1 responses before studying list 2 would promote study-phase retrieval of list 1 pairs in list 2, leading to greater memory interdependence on the final test.

Results

The interpolated task phase included either interpolated testing with and without feedback (experiments 1 and 2) or interpolated testing without feedback followed by changed responses or no changed responses in list 2 (experiment 3). In all three experiments, participants ($N = 75$ [experiments 1 and 2], $N = 68$ [experiment 3]) recalled about half of the list 1 responses on average (Fig. 2A), and recall was not significantly different between conditions [largest $\chi^2(1) = 0.92$, $P = 0.34$]. This performance level was ideal for assessing subsequent memory effects conditioned on interpolated retrieval success that we report later because items could be distributed somewhat evenly between cells.

Experiments 1 and 2

Experiments 1 and 2 used the same design and similar procedures and therefore are reported together. Experiment 1 established interpolated retrieval effects on existing memories and response interdependence with changed responses. Experiment 2 added an overt measure of change detection and recall of list 1 responses in list 2 to confirm differences in task effects on list 1 accessibility that we inferred from differences in list 1 accessibility on the final test in experiment 1. To provide an alternative response option when participants were asked whether responses had changed, we included an A–B, A–B repetition filler condition in these experiments. We did not have any theoretical interest in this condition. Starting with the final cued recall results, we first examined overall recall accuracy for existing memories based on the interpolated task types. Correct recalls occurred when participants reported the B response on the first cued recall prompt (for list 1) and the C response on the second cued recall prompt (for list 2). Intrusions from list 2 occurred when participants reported the C response on the first cued recall prompt. We assessed memory interdependence by conditioning list 1 recall on list 2 recall.

Figure 3A displays the correct list 1 recall. We examined interpolated task effects using separate models with the factor item type. Both models indicated a significant item type effect [smallest $\chi^2(4) = 536.77$, $P < 0.001$]. In experiment 1 (Fig. 3A, left panel), recall was significantly different for every pairwise comparison (smallest z ratio $= 3.31$, $P < 0.01$). In experiment 2 (Fig. 3A, middle panel), recall was significantly different for every pairwise comparison (smallest z ratio $= 3.81$, $P < 0.01$) except for the comparison of the A–B, A–B items and A–B, A–C items, including interpolated A–B testing with feedback (z ratio $= 1.60$, $P = 0.50$). This discrepancy in patterns between experiments likely occurred because list 2 items appeared longer in experiment 2 (8 sec) than in experiment 1 (6 sec). The extra time may have increased final B response recall more for A–B items that repeated across lists than for A–B, A–C items with interpolated tests and feedback. Moreover, these findings show retroactive facilitation in interpolated test conditions that was enhanced by feedback. The facilitation differences indicated that feedback increased list 1 response accessibility; this was confirmed by parallel patterns of list 1 recall in list 2 in experiment 2 (Fig. 2B).

Figure 3B (left and middle panels) displays intrusions from list 2. Note that intrusions in the A–B, A–C conditions were responses that had appeared in list 2 and therefore were episodic memory errors. In contrast, intrusions in the other conditions were responses that had not appeared in list 2 and therefore were semantic

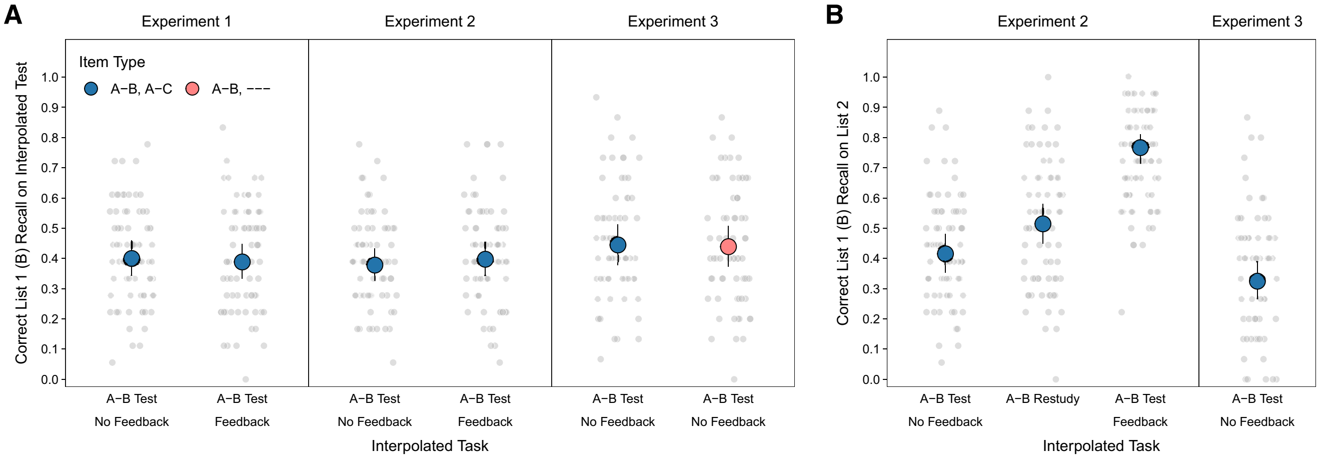

Figure 2. List 1 recall on interpolated tests (*A*) and in list 2 (*B*). The interpolated task labels on the *X*-axis of each panel indicate what happened with A–B pairs in the interpolated phase before participants learned A–C pairs in list 2. The interpolated tasks comprised cued recall tests of B responses without feedback (A–B test, no feedback) or with feedback (A–B test, feedback), and restudy of complete A–B pairs (A–B restudy). The item type coloring indicates whether pairs appeared in an A–B, A–C condition that included the same cues (A) in each list and different responses in list 1 (B) and list 2 (C) or in an A–B, — condition that included list 1 A–B pairs that did not correspond to any list 2 pairs. The colored points are marginal means estimated from mixed-effect models, and error bars are the corresponding 95% confidence intervals. The gray points are individual participant probabilities.

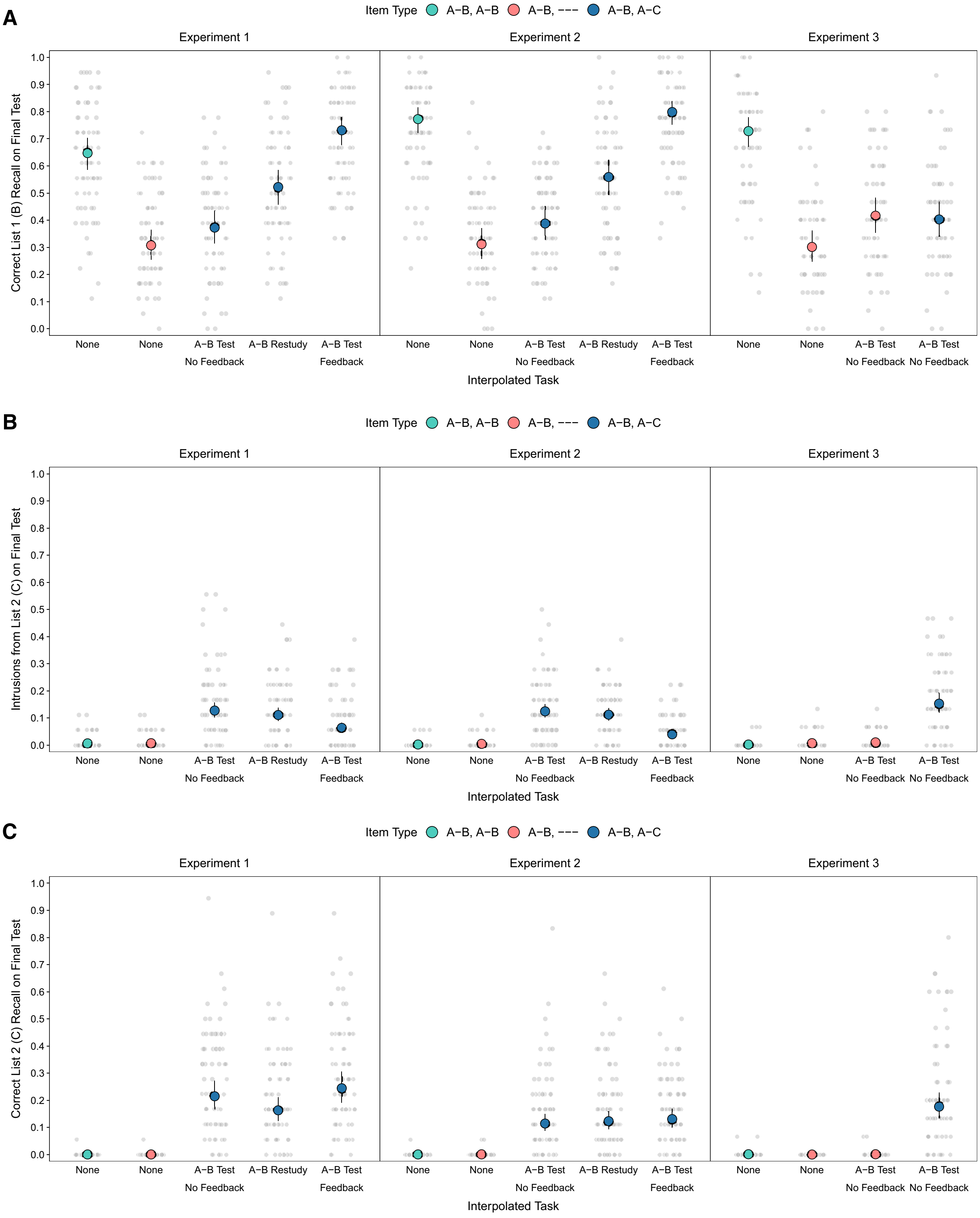

Figure 3. Correct list 1 recall (*A*), intrusions from list 2 (*B*), and correct list 2 recall (C) on the final test. The interpolated task labels on the *X*-axis of each panel indicate what happened with A–B pairs in the interpolated phase before participants learned A–C pairs in list 2. The interpolated tasks comprised cued recall tests of B responses without feedback (A–B test, no feedback), restudy of complete A–B pairs (A–B restudy), and cued recall tests of B responses with feedback (A–B test, feedback). The item type coloring indicates whether pairs appeared in an A–B, A–B condition that included the same cues (A) and responses (B) in lists 1 and 2; in an A–B, — condition that included list 1 A–B pairs that did not correspond to any list 2 pairs; or in an A–B, A–C condition that included the same cues (A) in each list and different responses in list 1 (B) and list 2 (C). Note that all C responses produced in the A–B, A–B and A–B, — conditions are extraexperimental intrusions. Those C response probabilities in *B* and *C* therefore show baseline estimates of participants guessing what would have been C responses for those items had they appeared in an A–B, A–C condition. The colored points are marginal means estimated from mixed-effect models, and error bars are the corresponding 95% confidence intervals. The gray points are individual participant probabilities.

memory errors (i.e., extraexperimental intrusions). The latter were included in the model as baseline measures of how often participants self-generated responses that would have appeared in list 2 in the A–B, A–C conditions. We examined the effects of the interpolated task manipulation using separate models with the factor item type. Both models indicated a significant item type effect [smallest $\chi^2(4) = 213.86$, $P < 0.001$]. For the A–B, A–C items, intrusion rates were not significantly different between conditions, including interpolated A–B testing without feedback and restudy [largest z ratio = 1.40, $P = 0.63$], but were significantly lower in the interpolated testing with feedback than the interpolated restudy condition (smallest z ratio = 4.74, $P < 0.001$). These results showed that list 2 responses interfered the least when interpolated retrieval of list 1 responses included corrective feedback.

Figure 3C (left and middle panels) displays correct list 2 recall. We examined interpolated task effects using the same model as above. As for intrusions from list 2 above, list 2 recall in conditions other than the A–B, A–C conditions were extraexperimental intrusions showing rates of self-generated responses that would have appeared had those items been in the A–B, A–C conditions. Both models indicated a significant item type effect [smallest $\chi^2(4) = 93.86$, $P < 0.001$]. In experiment 1, list 2 recall was significantly higher in both interpolated test conditions than in the interpolated restudy condition (smallest z ratio = 3.41, $P < 0.01$) and was not significantly different between interpolated test conditions (z ratio = 1.76, $P = 0.40$). However, in experiment 2, there were no significant differences across the A–B, A–C conditions (largest z ratio = 1.39, $P = 0.63$); the significant effect was driven by higher recall in the A–B, A–C than the other conditions (smallest z ratio = 5.92, $P < 0.001$). These results show that interpolated retrieval led to a forward testing benefit on list 2 recall, but this did not replicate when participants divided their attention during list 2 study in order to detect changes and recall list 1 responses.

Studies using A–B, A–C tasks have shown that conditions that promote list 1 recall during the list 2 study phase increases subsequent recall of both responses (e.g., see Postman and Gray 1977; Negley et al. 2018). Here, study-phase retrievals of list 1 responses during list 2 should have promoted memory interdependence, leading list 1 recall to be higher when list 2 responses were also recalled. However, it is important to note that such conditional list 1 recall is not a pure measure of memory interdependence because both responses sometimes, but very rarely, are recalled together when list 1 recall had not occurred during list 2. Also, intrusions from list 2 during the final recall of list 1 responses necessarily reduce the accuracy of list 2 responses. However, we consider this a reflection of the interference experienced when memory interdependence was not established. When list 2 responses are recalled, the absence of memory interdependence could still lead to omission and extraexperimental error rates comparable with the total rates of other error types when list 2 responses are recalled correctly. Moreover, if interpolated retrieval increased list 1 response accessibility, then responses from both lists should be recalled together on the final test more often for items recalled correctly than incorrectly during the interpolated phase even when corrective feedback followed. Finally, if interpolated retrieval enhanced final recall partly by promoting cross-episode associations that enhanced memory for items and their contexts, then final list 1 recall for items recalled in the interpolated phase should be greater when list 2 responses are recalled. We tested these predictions by examining memory interdependence for the conditions with interpolated tests.

We first assessed whether there was evidence for memory interdependence by examining whether list 1 recall depended on list 2 recall (Fig. 4, left and middle panels) using separate 2(item type) × 2(list 2 recall) models. The models indicated significant list 2 recall effects [smallest $\chi^2(1) = 32.35$, $P < 0.001$], showing higher list 1 recall when list 2 responses were recalled than when they were not. This effect was qualified by a significant interaction in experiment 1 [$\chi^2(1) = 5.45$, $P = 0.02$], showing that list 1 recall depended more on list 2 recall for interpolated retrieval without feedback; this effect was not significant in experiment 2 [$\chi^2(1) = 0.02$, $P = 0.90$], but the pattern paralleled experiment 1. The difference between experiments may have reflected the increased list 1 recall from overt list 1 retrieval attempts on list 2. These results suggest that study-phase retrievals promoted memory interdependence.

We next assessed the role of interpolated retrieval in promoting the memory interdependence shown in the previous conditional analyses (Fig. 5, left and middle panels). We first examined the role of interpolated retrieval using separate 2(interpolated recall) × 2(item type) models. The models indicated significant interpolated recall and item type effects [smallest $\chi^2(1) = 230.52$, $P < 0.001$] that were qualified by significant interactions [smallest

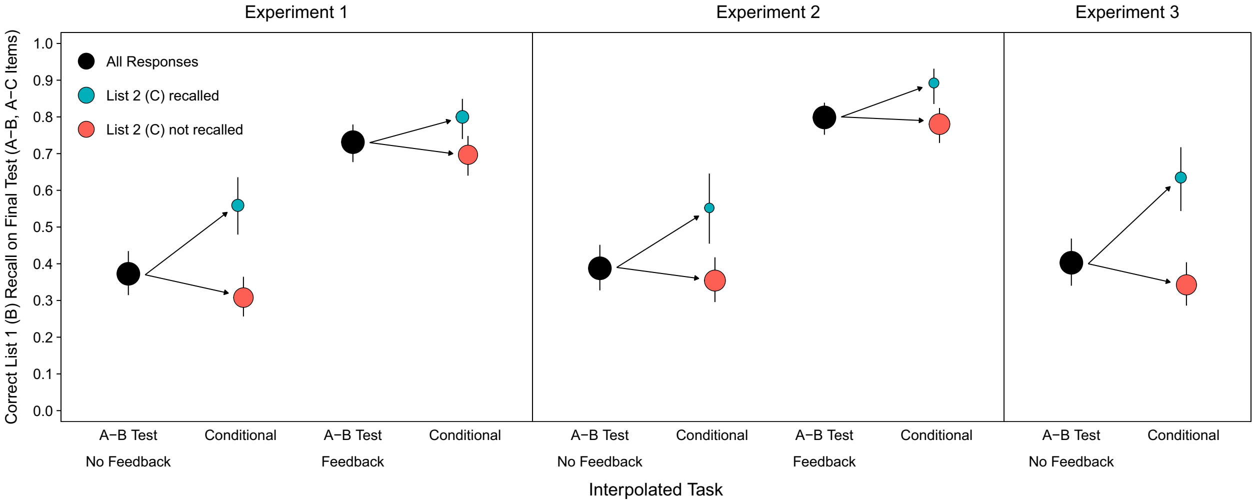

Figure 4. Final test list 1 recall conditioned on list 2 recall for A–B, A–C item types for which the interpolated tasks included tests of A–B pairs. The interpolated task labels on the *X*-axis of each panel indicate what happened with A–B pairs in the interpolated phase before participants learned A–C pairs in list 2. The probabilities for "conditional" items were conditioned on whether the C response was recalled correctly. The point heights are the marginal means estimated from mixed-effect models, and error bars are 95% confidence intervals. The conditional point sizes indicate observation differences.

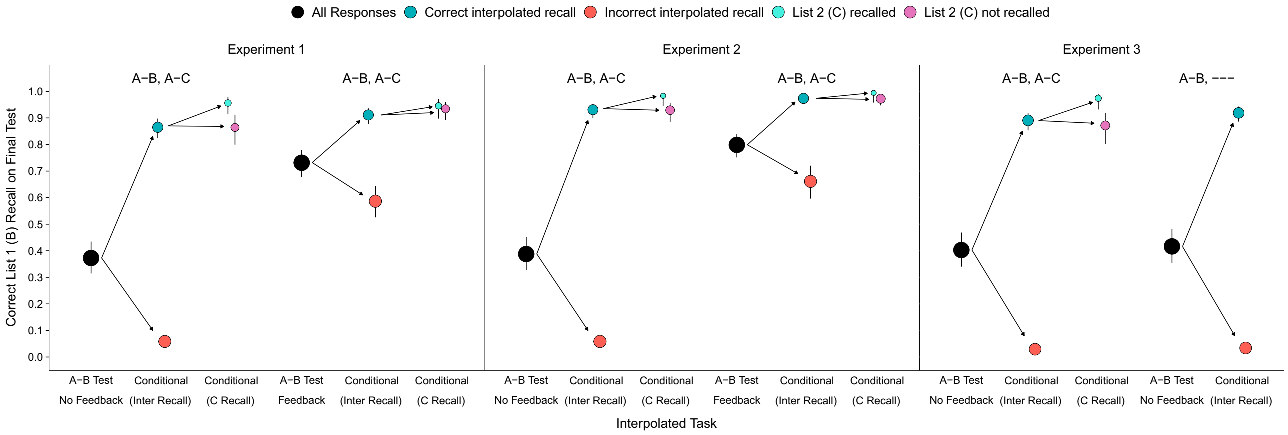

Figure 5. Final test list 1 recall, conditioned on recall in the interpolated phase and list 2. The interpolated task labels on the *X*-axis of each panel indicate what happened with A–B pairs in the interpolated phase before participants learned A–C pairs in list 2. The probabilities for "conditional" items were conditioned on whether the B response was recalled during the interpolated phase (interrecall) and whether the C response was recalled correctly (C recall) following correct interpolated B response recall. The point heights are the marginal means estimated from mixed-effect models, and error bars are 95% confidence intervals. The conditional point sizes indicate observation differences.

$\chi^2(1) = 49.64$, $P < 0.001$]. List 1 recall was significantly higher following correct than incorrect interpolated recall to a greater extent for interpolated testing without feedback (smallest z ratio = 2.42, $P = 0.02$) than with feedback (smallest z ratio = 18.09, $P < 0.001$). We then examined memory interdependence for successful interpolated retrievals, assuming that these best enabled cross-episode encoding because previously retrieved list 1 responses were most accessible in list 2 (see below for confirmation). Separate 2(item type) × 2(list 2 recall) models indicated significant item type and list 2 recall effects [smallest $\chi^2(1) = 6.10$, $P = 0.01$]. The interaction was significant in experiment 1 [$\chi^2(1) = 4.21$, $P = 0.04$] but not in experiment 2 [$\chi^2(1) = 0.01$, $P = 0.93$], but the pairwise comparisons yielded parallel findings: List 1 recall was significantly higher when list 2 responses were also recalled when feedback was not provided after interpolated retrievals (smallest z ratio = 2.38, $P = 0.02$), whereas list 1 recall was not significantly different based on list 2 recall when feedback was provided (largest z ratio = 1.53, $P = 0.13$). These results replicate the finding that testing effects depend on retrieval success (for review, see Rowland 2014) and suggest that interpolated retrieval enhanced final list 1 recall via retrieval practice per se and by promoting memory interdependence.

Finally, we verified that interpolated retrieval increased list 1 response accessibility during list 2 by conditioning list 1 recall during list 2 on interpolated retrieval success in experiment 2 (Fig. 6, left panel). A 2(item type) × 2(interpolated recall) model indicated a significant interpolated recall effect [$\chi^2(1) = 490.16$, $P < 0.001$] and a significant interaction [$\chi^2(1) = 79.19$, $P < 0.001$]. List 1 recall was higher following correct than incorrect interpolated retrievals to a greater extent without feedback, suggesting that interpolated retrieval promoted list 1 recall on list 2 study beyond re-exposure from feedback.

In sum, experiments 1 and 2 showed that interpolated retrieval improved memory for list 1 responses, resulting in retroactive facilitation that was greater when feedback appeared after interpolated retrieval attempts. Two key findings are compatible with the proposal that successful interpolated retrieval promotes integrative encoding that facilitates subsequent memory. Final list 1 recall was better when list 2 responses were also recalled, and final list 1 recall after correct interpolated retrieval was better when list 2 responses were also recalled and feedback was not provided after interpolated retrieval. These results suggest that retrieval-enhanced accessibility of existing memories increased the extent to which list 2 items with shared and changed features cued retrieval of list 1 memories. Indeed, experiment 2 showed that list 1 accessibility during list 2, which is required for episodes to be retroactively integrated, was higher after correct than incorrect interpolated retrievals.

Experiment 3

One could argue that the interpolated retrieval benefits on final list 1 recall solely reflect retrieval practice effects. Indeed, all the accounts considered here predict that interpolated retrieval should enhance list 1 recall in both interpolated test conditions compared with the control condition in which only A–B pairs appeared in list 1. To evaluate the retrieval practice account, we included a condition in which list 1 items were tested without feedback in the interpolated phase and without corresponding A–C items in list 2 (condition A–B, —). The retrieval practice account predicts worse list 1 recall when interpolated retrieval practice is followed by changed responses in list 2 than when no changed responses appear because response competition would counteract retrieval practice benefits. In contrast, the differentiation and integration accounts both predict comparable performance regardless of whether A–C items appear in list 2 because those accounts assume that interpolated retrieval mitigates the negative effects of response competition. Importantly, the condition including A–B, A–C pairs and interpolated testing without feedback provides an important test of the integration account because that account predicts memory interdependence of the sort that was observed in experiments 1 and 2.

The task in experiment 3 maintained several elements of the task in experiment 2 (see Fig. 1). However, we reduced the conditions to focus on the comparison of interpolated testing with and without subsequent changed responses. The design therefore included those two conditions along with the previous control and repeated filler conditions. We also reduced the items per cell to shorten the task because the pandemic required us to develop a task that could be administered remotely. This method may have led to lower overall performance, as seen, for example, in list 1 accessibility during list 2 in the A–B, A–C, with the interpolated testing condition being lower than the previous experiments (Fig. 2). Nevertheless, the recall patterns reported below are sensible given the findings in experiments 1 and 2.

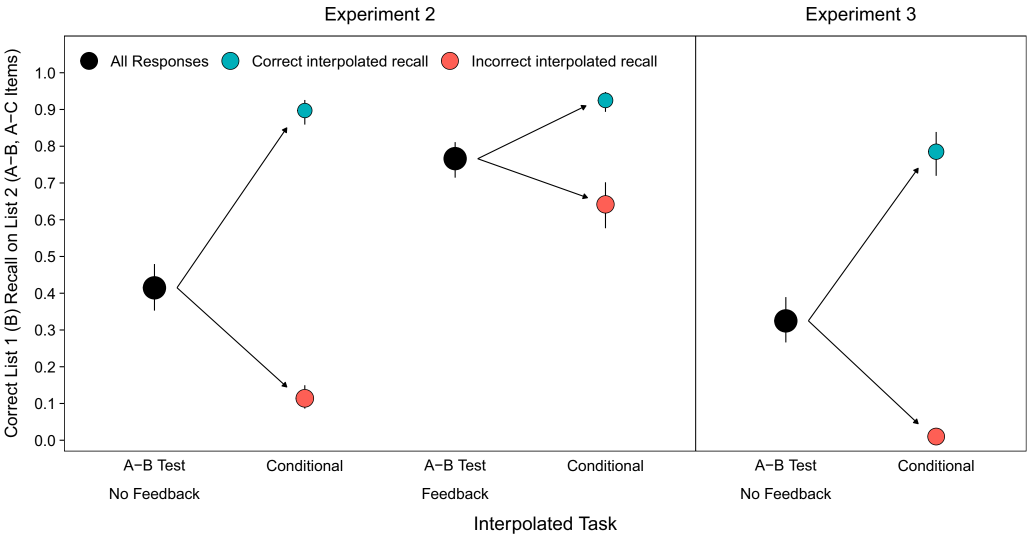

Figure 6. List 1 recall on list 2, conditioned on recall during the interpolated phase for A–B, A–C item types. The interpolated task labels on the *X*-axis of each panel indicate what happened with A–B pairs in the interpolated phase before participants learned A–C pairs in list 2. The probabilities for items labeled "conditional" were conditioned on whether the B response was recalled during the interpolated phase. The point heights are the marginal means estimated from mixed-effect models, and error bars are 95% confidence intervals. The conditional point sizes indicate observation differences.

Figure 3A (right panel) displays correct list 1 recall on the final test. A model with the factor item type indicated a significant effect [$\chi^2(3) = 330.14$, $P < 0.001$], showing that recall was significantly different for every pairwise comparison (smallest *z* ratio = 4.48, $P < 0.001$) except for the interpolated test conditions (*z* ratio = 0.58, $P = 0.94$). List 1 recall was significantly higher in the interpolated test conditions than the A–B, — control condition, showing that retrieval practice led to comparable retroactive facilitation regardless of whether a changed response subsequently appeared. Figure 3B (right panel) displays intrusions from list 2. A model with the factor item type indicated a significant effect [$\chi^2(3) = 224.28$, $P < 0.001$], showing significantly more episodic memory errors in the A–B, A–C with the interpolated testing condition than semantic memory errors in the other conditions (smallest *z* ratio = 7.57, $P < 0.001$). Figure 3C (right panel) displays correct list 2 recall. A model with the factor item type indicated a significant effect [$\chi^2(3) = 110.31$, $P < 0.001$], showing significantly more correct recalls in the A–B, A–C with the interpolated testing condition than guesses in the other conditions (smallest *z* ratio = 7.50, $P < 0.001$).

As in experiments 1 and 2, we first sought evidence for memory interdependence by conditioning list 1 recall on list 2 recall (Fig. 4, right panel) using a model with list 2 recall as a factor. The model indicated a significant effect [$\chi^2(1) = 42.35$, $P < 0.001$], showing higher list 1 recall when list 2 responses were also recalled. Also, as in the prior experiments, we conditioned list 1 recall on interpolated recall in both interpolated test conditions (Fig. 5, right panel) to assess the contribution of retrieval practice to final list 1 recall. A 2(interpolated recall) × 2(item type) model indicated a significant interpolated recall effect [$\chi^2(1) = 597.61$, $P < 0.001$] and no other significant effects [largest $\chi^2(1) = 2.19$, $P = 0.14$], showing that list 1 recall was significantly higher following correct than incorrect interpolated retrieval. We also verified that list 1 responses were more accessible during list 2 study following correct than incorrect interpolated retrieval [$\chi^2(1) = 199.83$, $P < 0.001$] (Fig. 6, right panel). Finally, the comparable interpolated retrieval benefits with and without changed responses suggested that retrieval practice increased list 1 response accessibility but also counteracted interference. A follow-up analysis of only the A–B, A–C with interpolated testing showed that list 1 recall was significantly higher when list 2 recall was correct rather than incorrect (*z* ratio = 3.38, $P < 0.001$). Collectively, these results show that response competition after interpolated testing retroactively interfered with list 1 recall, but this interference was perfectly offset by the facilitation in list 1 recall associated with list 2 recall.

In sum, experiment 3 replicated key patterns from experiment 2. Interpolated retrieval facilitated final list 1 recall, final list 1 recall was better when list 2 responses were also recalled, and list 1 recall following correct interpolated retrievals was also better when list 2 responses were recalled. Experiment 3 uniquely showed that final list 1 recall following successful interpolated retrieval was comparable regardless of whether competing responses appeared in list 2, providing evidence against a pure retrieval practice account of the interpolated testing effects observed in the present study. Together with the results replicating the prior experiments, this finding further supports the proposal that interpolated retrieval promotes the encoding of cross-episode associations that enhance memory for items and their contexts.

Discussion

By promoting the act of retrieval, interpolated testing has been shown to improve existing memories under conditions in which response competition is present. Two theories accounting for these benefits have often been considered. The context differentiation view proposes that testing counteracts response competition by associating context more uniquely with individual events. In contrast, the integrative encoding view proposes that testing reduces interference by making it more likely that people will incorporate existing memories with new learning when the two are coactivated. Note that the dichotomy between these accounts may be overstated, as integrative encoding involves representing items in association with each other and with their separate contexts. The present study used A–B, A–C tasks to test the integrative encoding hypothesis that successful retrieval of list 1 (B) responses should make them more accessible when studying list 2 pairs with changed (C) responses, leading to interdependence (BC) that retroactively enhances existing memories. Supporting this account, list 1 responses that were correctly recalled on interpolated test trials were more often recalled on list 2. These instances of list 1 recall on list 2 were associated with retroactive facilitation and interdependence in the recall of responses from both lists on the final test. These testing benefits were attributable to interpolated retrieval instead of re-exposure, as response interdependence was greater for correct than incorrect interpolated retrievals even when the latter included feedback. These findings are incompatible with a strict context differentiation prediction that coactivating responses during encoding and response interdependence at test should be associated with impaired memory accuracy. They are more compatible with the integrative encoding view that such coactivation and interdependence should be associated with improved memory for both item content and associated context.

The present findings illuminate theoretical issues about interpolated retrieval effects when response competition is present. A foundational study reporting test-enhanced memory in A–B, A–C tasks found this benefit regardless of whether response changes were manipulated between or within subjects and attributed those effects to improved list differentiation (Tulving and Watkins 1974). However, studies using similar tasks without interpolated testing suggested that awareness of response changes could lead to retroactive facilitation by inducing list 1 recall on list 2 (Bruce and Weaver 1973; Robbins and Bray 1974), which undermines list differentiation. These findings suggest that interpolated retrieval leads to more recall of list 1 pairs on list 2. Our earlier findings support this prediction in showing that interpolated testing eliminated proactive interference partly by improving recollection that responses had changed (Wahlheim 2015) and memory for the list 1 responses themselves (Kemp et al. 2023), which were assumed to be the downstream consequences of study-phase retrievals and detection of changed responses. The present study directly examined the roles of such processes on list 2 by including overt measures upon which to condition final recall performance. The results showing increased response interdependence for items recalled on interpolated test trials implicates a role for integrative encoding.

The present findings are generally consistent with a theoretical framework of episodic memory updating that proposes key roles for detecting changes when they occur and later remembering that such changes had been detected in subsequent memory for items and their contexts. The “memory for change” framework (for review, see Wahlheim et al. 2021) proposes that features of current events can trigger retrieval of existing memories with shared features, thus providing opportunities for integrative encoding that enhances memory when people recollect event changes and the relative temporal association between them (cf. Hintzman 2011). Here and in similar tasks, target list recall was enhanced when participants could remember that responses had changed and the competing response (for review, see Wahlheim et al. 2021). In A–B, A–C tasks used to examine retroactive effects of memory, such change recollection was more likely when participants were given more time to study list 2 (Negley et al. 2018; Garlitch and Wahlheim 2020) and was associated with better memory for list 1 responses relative to when changes were not remembered. The present study adds to these findings in showing that interpolated list 1 recall led to better subsequent recall of responses from both lists that required recollection of responses having changed. Collectively, these results show that providing more opportunities to retrieve features of existing memories to detect changes in sensory inputs can promote better recall of both existing memories and recent events with shared and changed features.

Related work has shown that retroactive interference varies when interpolated tasks are manipulated between participants in an A–B, A–C task with phases separated by 48 h (Scully and Hupbach 2020). Interpolated cued recall (without feedback) and restudy both led to better list 1 recall and fewer intrusions from list 2 than interpolated list 1 cue ratings and a distractor activity. These results suggested that cuing retrieval of existing memories reduced competition from new learning. Although their study used A–B, A–C tasks as in the present investigation, the different approaches complicated comparisons. One inconsistency was that they observed better final list 1 recall in the interpolated test than in the restudy condition, whereas we found the opposite. They observed this because retrieval practice without feedback often enhances memory on delayed but not immediate tests (e.g., see Roediger and Karpicke 2006). Interpolated recall was also facilitated in their study by criterion learning before interpolated testing, whereas our task produced intermediate interpolated recall necessary for conditional analyses. Finally, their analytic approach precluded inferences about differences in memory interdependence.

Similar research has also been conducted in the context of postevent misinformation effects using A–B, A–C tasks with videos, pictures, and narratives of everyday events as stimuli. Misinformation effects occur when misleading information interferes retroactively with memory for original event details (for review, see Loftus 2005). As in A–B, A–C tasks, detecting discrepancies between original and misleading details can enhance existing memories (Tousignant et al. 1986), especially for people who detect and later recollect more discrepancies (Putnam et al. 2017). Counterintuitively, testing event details immediately after events can increase misinformation susceptibility (for review, see Chan et al. 2017). However, other work has shown that such retrieval-enhanced suggestibility occurred only when discrepancies were not remembered and did not occur when discrepancies were overtly detected (Butler and Loftus 2018). The latter finding suggests that seeking discrepancies can increase study-phase retrievals, providing more opportunities for integrative encoding. Others have noted that retrieval-enhanced suggestibility occurs when discrepancies are interpreted as corrective feedback after incorrect retrieval attempts, leading original event details to be rejected partly because the recognition procedure led participants to mistakenly endorse the discrepancy (Rindal et al. 2016).

Taken with the present results, the various interpolated retrieval effects in misinformation tasks suggest that postevent retrieval may diminish overall suggestibility. However, these benefits require that the task promotes retrieval of original details when detecting discrepancies for nearly every event. They also require that participants understand that discrepancies are not corrective feedback and that those discrepancies should be rejected on a later test. Future studies should thus consider interactions among original event detail memorability, interpolated retrieval cue strength, and how test instructions ensure understanding that discrepancies are incorrect responses. The latter may be addressed by using variants of the present test procedure that constrain participants’ retrieval to specific sources using test prompts that specify the modality in which information had appeared and its objective accuracy.

Neuroimaging studies of reactivation effects in A–B, A–C tasks provide converging evidence for the proposal that study-phase retrievals enable memory interdependence. Prior work has shown that univariate activation in particular hippocampus and frontostriatal regions during A–C learning is associated with better memory for A–B pairs, suggesting that reactivation protected existing memories against retroactive interference (Kuhl et al. 2010). Related work using pattern classification and A–B retrieval cues before A–C trials to examine the interaction of reactivation and integrative encoding suggests that reactivation states in the medial prefrontal cortex contribute to integrative encoding associated with reduced interference (Richter et al. 2016; Chanales et al. 2019). Converging results have been found in associative inference tasks in which participants learn changed pairs (A–B, B–C) and later infer A–C links (for review, see Schlichting and Preston 2015). Most relevant, after learning A–B pairs, neural reactivation during rest before B–C learning has been shown to predict more neural engagement in category-selective visual regions during B–C learning that supported reinstatement of A–B picture word pairs and facilitated B–C encoding (Schlichting and Preston 2014). The results were interpreted as showing that spontaneous reactivation strengthened A–B memories, increasing their accessibility during B–C study. This is compatible with the findings here that correct interpolated A–B recalls were more accessible during A–C study, which was associated with memory interdependence, reflecting the encoding of cross-episode associations. More generally, these findings are reminiscent of work showing that semantic associations promote feature binding that occurs in the hippocampus (Antony et al. 2022).

In conclusion, three experiments showed that interpolated retrieval success promoted memory interdependence that was associated with retroactive facilitation. These experiments made unique methodological advances by combining analyses of interpolated retrieval success at the trial level with overt measures of list 1 accessibility on list 2 and subsequent memory interdependence. The present findings are incompatible with the original account that interpolated testing in A–B, A–C tasks enhanced memory by promoting list differentiation. In contrast, the findings here are compatible with extant behavioral and neural evidence for integrative encoding across a wide variety of tasks ranging in their likeness to everyday situations. Adopting aspects of the current approach across literatures promises to aid the establishment of a unified account of change-cued reactivation effects on existing memories.

Materials and Methods

We report here how we determined sample sizes, all data exclusions, all manipulations, and all measures (Simmons et al. 2012). The current study was approved by the Institutional Review Board at the University of North Carolina at Greensboro (UNCG). The materials, data, and analysis code are available on the Open Science Framework (OSF; https://osf.io/w2au6).

Experiment 1

Participants

Participants were UNCG students who received partial course credit as compensation. We planned to recruit as many people as possible in one semester with the goals of having usable data from at least 60 participants and a final sample size that was a multiple of five (the number of experimental formats). The final sample included 75 participants (50 women and 25 men) ages 18–24 yr ($M = 18.57$ yr, SD = 1.06 yr). A sensitivity analysis was conducted to determine the smallest effect size that could be detected for the difference in final list 1 recall depending on list 2 recall for items recalled in the interpolated phase in the A–B, A–C condition that did not include feedback. We targeted this effect for sensitivity analysis because it has the fewest observations and provides a critical test of the main hypothesis. The analysis was conducted using functions from the simr package v.1.0.5 (Green and MacLeod 2016), with 500 simulations and the P-value method set to "lrt" (likelihood ratio test). An odds ratio (OR) as small as 2.49 could be detected with power = 80%, $\alpha = 0.05$, and $N = 75$. A post-hoc power curve (Supplemental Fig. S1) indicated that $N = 40$ was sufficient to detect the observed OR = 3.43 with power = 80% and $\alpha = 0.05$.

Design and materials

The experiment used a within-subject manipulation of item type that included A–B, A–B items; A–B, — items; and three types of A–B, A–C items that had different interpolated tasks. For the A–B, A–C items, the interpolated tasks were A–B tests without feedback, A–B restudy, and A–B tests with feedback. The materials comprised 95 three-word sets that included a cue word (e.g., clever) with two response words (e.g., wise and trick) taken from Wahlheim (2015). The associative strengths between words in each set were indexed by the free association norms (http://w3.usf.edu/FreeAssociation). The associative strengths between cues and responses were low on average (forward: $M = 0.04$, SD = 0.02, range 0.01–0.10; backward: $M = 0.02$, SD = 0.03, range 0.00–0.10). The responses in each word set were not associated. Of the 95 sets, five served as primacy and recency buffers in lists 1 and 2 and as practice test items. The remaining 90 sets served as critical items in all phases and were further divided into five groups of 18. The groups were rotated through each within-subject condition such that each appeared equally often across participants. For A–B, A–C items, the assignment of target (B) and competing (C) responses was not counterbalanced. This arrangement produced five experimental formats. Each buffer set was assigned to an item type condition and was not counterbalanced.

Figure 1 displays a schematic of the procedure. There were four phases: list 1, the interpolated tasks, list 2, and a final cued recall test. List 1 included 90 cue–response word pairs across all conditions. The interpolated phase included 54 items distributed evenly across the three A–B, A–C conditions. The 18 items from A–B restudy trials were cue–response word pairs, the 18 items from A–B tests without feedback were only cues, and the 18 A–B tests with feedback were cues and responses that appeared asynchronously. List 2 included 72 cue–response word pairs distributed evenly across conditions, excluding the A–B, — condition. The final cued recall test phase included 90 cue words distributed evenly across all conditions.

Procedure

Participants were tested individually. The stimuli were presented on computers using E-Prime 2.0 software (https://support.pstnet.com). In all phases of the experiment, stimuli appeared in lowercase white font against a black background. In every phase, stimuli appeared in a fixed random order with the restriction that no more than three items from the same condition appeared consecutively. The average list position was equated across conditions to control for serial position effects.

On list 1, word pairs (e.g., clever–wise, coffee–bean, strong–will, etc.) appeared for 4 sec each followed by a 0.5-sec interstimulus interval (ISI), during which the screen was blank. Participants were instructed to read the pairs aloud and to study them for an upcoming memory test. Next, participants completed the interpolated task phase, where they were told to expect three types of tasks. All items appeared for 6 sec each and were followed by a 0.5-sec ISI. For A–B restudy trials, word pairs appeared exactly as in list 1 (e.g., clever and wise). Participants were told to read them aloud and study them for a test. For A–B tests without feedback, items appeared as the cue word with a question mark (e.g., coffee — ?). Participants were instructed to read the cue and recall the list 1 response aloud while the cue was on the screen (e.g., coffee and bean). For A–B test trials with feedback, items appeared as the cue paired with a question mark for 4 sec (e.g., strong — ?). Participants were instructed to read the cue and recall the list 1 response aloud during that time (e.g., strong and will). After 4 sec, cue–question mark pairs remained on the screen, and the response from list 1 (e.g., will) appeared below in lowercase green font for 2 sec. Participants were instructed to read the feedback silently. All responses were recorded by an experimenter. Participants were instructed to pass when they could not remember responses, as opposed to guessing, to limit extraexperimental intrusions on the final cued recall test.

On list 2, word pairs appeared for 4 sec each and were followed by a 0.5-sec ISI. Participants were again instructed to read them aloud and study them for a test. Participants then completed the final cued recall test. On each trial, a cue–question mark pair appeared, and participants were told first to type the response that appeared with the cue in list 1. They were then told to indicate whether the response that was paired with the cue changed from list 1 to list 2. Participants pressed the "1" key to indicate that the response had changed and the "0" key to indicate that the response had not changed. When participants indicated that the response had changed, they were told to recall the list 2 response by typing it onto the screen. When they indicated that the response did not change, the program advanced to the next item.

Experiment 2

Participants

Participants were UNCG students who received partial course credit as compensation. Our recruitment plan paralleled experiment 1. We tested 80 people and excluded five who did not follow instructions, resulting in a final sample of 75 participants (50 women and 25 men), ages 18–32 yr ($M = 19.45$ yr, SD = 2.50 yr). The participants who did not follow instructions were identified based on their behavior during the experiment. These participants omitted responses on a substantial number of trials in one or more phases. A sensitivity analysis conducted as in experiment 1 indicated that an OR as small as 4.20 could be detected with power = 80%, $\alpha = 0.05$, and $N = 75$. A post-hoc power curve (Supplemental Fig. S2) determined that $N = 70$ was sufficient to detect the observed OR = 4.45 with power = 80% and $\alpha = 0.05$.

Design, materials, and procedure

The design, materials, and procedure were the same as for experiment 1, except in the procedure, the list 2 study phase included overt measures of change detection and list 1 recall. Participants were told that while list 2 word pairs were on the screen, they would be asked to identify pairs with the same left member and changed right member compared with pairs from list 1. List 2 pairs appeared for a total of 8 sec each followed by a 0.5-sec ISI. Pairs appeared alone for the first 4 sec so participants could encode before overtly responding. During the next 4 sec, the prompt "changed (1)" appeared below the pair. Participants were told to press the "1" key only when a changed pair appeared. Each pair disappeared after 8 sec had elapsed. For pairs identified as changed, participants attempted to recall the list 1 response by typing it onto the screen. For pairs not identified as changed, the program automatically advanced to the next item.

Experiment 3

Participants

We planned to test a number of UNCG students comparable with the prior experiments by recruiting as many participants as possible in one semester, stopping at a multiple of four (the number of experimental formats). The final sample included 68 participants (48 women and 20 men), ages 18–30 yr ($M = 20.03$ yr, SD = 2.64 yr), who received partial course credit as compensation. A sensitivity analysis conducted as before indicated that an OR as small as 3.45 could be detected with power = 80%, $\alpha = 0.05$, and $N = 68$. A post-hoc power curve (Supplemental Fig. S3) determined that $N = 32$ was sufficient to detect the observed OR = 5.51 with power = 80% and $\alpha = 0.05$.

Design, materials, and procedure

The design, materials, and procedure included many elements of experiment 2 with some differences. The design of experiment 3 included a within-subject manipulation of item type, including A–B, A–B items; A–B, — items; A–B, — items with interpolated testing and no feedback; and A–B, A–C items with interpolated testing and no feedback. All the conditions were the same as in the prior experiments, except for the A–B, — items with interpolated testing. That condition was the same as the A–B, A–C condition with interpolated testing, except that it did not include competing A–C items in list 2. Each condition included 15 critical items (60 total). The associative strengths between cues and responses were low on average (forward: $M = 0.04$, SD = 0.02, range 0.01–0.10; backward: $M = 0.02$, SD = 0.03, range 0.00–0.18). List 1 included 60 items, the interpolated task phase included 30 items, list 2 included 30 items, and the final cued recall test included 60 items. We reduced the number of items and conditions to promote sustained attention to the task. This was necessary because the COVID-19 pandemic compelled us to develop a virtual synchronous testing protocol that took longer to administer and could be completed outside the laboratory. Participants downloaded and deployed a program in E-Prime Go software (https://support.pstnet.com). Using the Zoom videoconferencing application, a research assistant explained the tasks and monitored participant behavior. The data files automatically uploaded to a server hosted by E-Prime when the experiment ended.

Statistical methods

All data preprocessing and statistical tests were performed using R software v.4.1.0 (R Core Team 2021). We wrangled and visualized the data using functions from the tidyverse package v.1.3.1 (Wickham et al. 2019). We examined the effects of the experimental manipulation using generalized linear mixed-effect models fitted with functions from the lme4 package v.1.1.27.1 (Bates et al. 2015). These models included random intercept effects of participants and items to maximize power. They also included fixed effects of within-subject conditions. The specifications for all models are on the OSF website (https://osf.io/w2au6). After fitting the models, we performed Wald's χ^2 hypothesis tests using the Anova function of the car package v.3.0.10 (http://socserv.socsci.mcmaster.ca/jfox/Books/Companion) and pairwise comparisons using the emmeans function from the emmeans package v.1.6.1 (https://CRAN-Rproject.org/package=emmeans) with the Tukey method to control for multiple comparisons. The significance level was $\alpha = 0.05$.

The materials, data, and analysis code for these experiments are available on the Open Science Framework (https://osf.io/w2au6).

Acknowledgments

We thank Brittani Marceno, Carson Peske, Cayla Kitts, Crystal Thinzar, Danni Brower, Desmond Vasquez, and Skye Harrelson for their assistance with data collection.

References

Abra JC. 1972. List differentiation and forgetting. In: *Human memory: Festschrift for Benton J. Underwood* (ed. Duncan CP, et al.), pp. 25–57. Appleton-Century-Crofts, New York.

Anderson MC, Neely JH. 1996. Interference and inhibition in memory retrieval. In *Memory* (ed. Bjork EL, Bjork RA), pp. 237–313. Academic Press, San Diego, CA.

Antony JW, Romero A, Vierra AH, Luenser RS, Hawkins RD, Bennion KA. 2022. Semantic relatedness retroactively boosts memory and promotes memory interdependence across episodes. *Elife* **11:** e72519. doi:10.7554/eLife.72519

Arkes HR, Lyons DJ. 1979. A mediational explanation of the priority effect. *J Verb Learn Verb Behav* **18:** 721–731. doi:10.1016/S0022-5371(79)90425-0

Barnes JM, Underwood BJ. 1959. 'Fate' of first-list associations in transfer theory. *J Exp Psychol* **58:** 97–105. doi:10.1037/h0047507

Bates D, Mächler M, Bolker B, Walker S. 2015. Fitting linear mixed-effects models using lme4. *J Stat Softw* **67:** 1–48. doi:10.18637/jss.v067.i01

Benjamin AS, Ross BH. 2010. The causes and consequences of reminding. In *Successful remembering and successful forgetting: a festschrift in honor of Robert A. Bjork* (ed. Benjamin AS), pp. 71–87. Psychology Press, New York.

Bower GH. 1972. Stimulus-sampling theory of encoding variability. In *Coding processes in human memory* (ed. Melton AW, Martin E), pp. 85–121. Wiley, Washington, DC.

Bramão I, Jiang J, Wagner AD, Johansson M. 2022. Encoding contexts are incidentally reinstated during competitive retrieval and track the temporal dynamics of memory interference. *Cereb Cortex* **32:** 5020–5035. doi:10.1093/cercor/bhab529

Bruce D, Weaver GE. 1973. Retroactive facilitation in short-term retention of minimally learned paired associates. *J Exp Psychol* **100:** 9–17. doi:10.1037/h0035488

Butler BJ, Loftus EF. 2018. Discrepancy detection in the retrieval-enhanced suggestibility paradigm. *Memory* **26:** 483–492. doi:10.1080/09658211.2017.1371193

Chan JCK, Manley KD, Lang K. 2017. Retrieval-enhanced suggestibility: a retrospective and a new investigation. *J Appl Res Mem Cogn* **6:** 213–229. doi:10.1016/j.jarmac.2017.07.003

Chan JCK, Manley KD, Davis SD, Szpunar KK. 2018a. Testing potentiates new learning across a retention interval and a lag: a strategy change perspective. *J Mem Lang* **102:** 83–96. doi:10.1016/j.jml.2018.05.007

Chan JCK, Meissner CA, Davis SD. 2018b. Retrieval potentiates new learning: a theoretical and meta-analytic review. *Psychol Bull* **144:** 1111–1146. doi:10.1037/bul0000166

Chanales AJH, Dudukovic NM, Richter FR, Kuhl BA. 2019. Interference between overlapping memories is predicted by neural states during learning. *Nature Commun* **10:** 5363. doi:10.1038/s41467-019-13377-x

Cox WR, Dobbelaar S, Meeter M, Kindt M, van Ast VA. 2021. Episodic memory enhancement versus impairment is determined by contextual similarity across events. *Proc Natl Acad Sci* **118:** e2101509118. doi:10.1073/pnas.2101509118

Darley CF, Murdock BB. 1971. Effects of prior free recall testing on final recall and recognition. *J Exp Psychol* **91:** 66–73. doi:10.1037/h0031836

Favila SE, Chanales AJ, Kuhl BA. 2016. Experience-dependent hippocampal pattern differentiation prevents interference during subsequent learning. *Nat Commun* **7:** 11066. doi:10.1038/ncomms11066

Garlitch SM, Wahlheim CN. 2020. The role of reminding in retroactive effects of memory for older and younger adults. *Psychol Aging* **35:** 697–709. doi:10.1037/pag0000427

Gershman SJ, Schapiro AC, Hupbach A, Norman KA. 2013. Neural context reinstatement predicts memory misattribution. *J Neurosci* **33:** 8590–8595. doi:10.1523/JNEUROSCI.0096-13.2013

Green P, MacLeod CJ. 2016. simr: an R package for power analysis of generalized linear mixed models by simulation. *Methods Ecol Evol* **7:** 493–498. doi:10.1111/2041-210X.12504

Henson RN, Shallice T, Josephs O, Dolan RJ. 2002. Functional magnetic resonance imaging of proactive interference during spoken cued recall. *Neuroimage* **17:** 543–558. doi:10.1006/nimg.2002.1229

Hintzman DL. 2011. Research strategy in the study of memory: fads, fallacies, and the search for the 'coordinates of truth'. *Perspect Psychol Sci* **6:** 253–271. doi:10.1177/1745691611406924

Horner AJ, Bisby JA, Bush D, Lin WJ, Burgess N. 2015. Evidence for holistic episodic recollection via hippocampal pattern completion. *Nat Commun* **6:** 7462. doi:10.1038/ncomms8462

Horner AJ, Burgess N. 2014. Pattern completion in multielement event engrams. *Curr Biol* **24:** 988–992. doi:10.1016/j.cub.2014.03.012

Howard MW, Kahana MJ. 2002. A distributed representation of temporal context. *J Math Psychol* **46:** 269–299.

Jacoby LL. 1974. The role of mental contiguity in memory: registration and retrieval effects. *J Verb Learn Verb Behav* **13:** 483–496. doi:10.1016/S0022-5371(74)80001-0

Jacoby LL, Wahlheim CN. 2013. On the importance of looking back: the role of recursive remindings in recency judgments and cued recall. *Mem Cognit* **41:** 625–637. doi:10.3758/s13421-013-0298-5

Jacoby LL, Wahlheim CN, Kelley CM. 2015. Memory consequences of looking back to notice change: retroactive and proactive facilitation. *J Exp Psychol Learn Mem Cogn* **41:** 1282–1297. doi:10.1037/xlm0000123

Jang Y, Huber DE. 2008. Context retrieval and context change in free recall: recalling from long-term memory drives list isolation. *J Exp Psychol Learn Mem Cogn* **34:** 112–127. doi:10.1037/0278-7393.34.1.112

Kemp PL, Loaiza VM, Wahlheim CN. 2023. Testing can enhance episodic memory updating in younger and older adults. *Psychol Aging* (in press).

Koen JD, Rugg MD. 2016. Memory reactivation predicts resistance to retroactive interference: evidence from multivariate classification and pattern similarity analyses. *J Neurosci* **36:** 4389–4399. doi:10.1523/JNEUROSCI.4099-15.2016

Kuhl BA, Shah AT, DuBrow S, Wagner AD. 2010. Resistance to forgetting associated with hippocampus-mediated reactivation during new learning. *Nat Neurosci* **13:** 501–506. doi:10.1038/nn.2498

Kumaran D, McClelland JL. 2012. Generalization through the recurrent interaction of episodic memories: a model of the hippocampal system. *Psychol Rev* **119:** 573–616. doi:10.1037/a0028681

Loftus EF. 2005. Planting misinformation in the human mind: a 30-year investigation of the malleability of memory. *Learn Mem* **12:** 361–366. doi:10.1101/lm.94705

Martin E. 1972. Stimulus encoding in learning and transfer. In *Coding processes in human memory* (ed. Melton AW, Martin E), pp. 85–121. Wiley, Washington, DC.

Mensink G-J, Raaijmakers JGW. 1988. A model for interference and forgetting. *Psychol Rev* **95:** 434–455. doi:10.1037/0033-295X.95.4.434

Müller GE, Pilzecker A. 1900. Experimentalle beitrage zur lehre com gedachtnis. *Z Psychol* **1:** 1–300.

Negley JH, Kelley CM, Jacoby LL. 2018. The importance of time to think back: the role of reminding in retroactive effects of memory. *J Exp Psychol Learn Mem Cogn* **44:** 1352–1364. doi:10.1037/xlm0000512

Ngo CT, Horner AJ, Newcombe NS, Olson IR. 2019. Development of holistic episodic recollection. *Psychol Sci* **30:** 1696–1706. doi:10.1177/0956797619879441

Ngo CT, Benear SL, Popal H, Olson IR, Newcombe NS. 2021. Contingency of semantic generalization on episodic specificity varies across development. *Curr Biol* **31:** 2690–2697. doi:10.1016/j.cub.2021.03.088

Osgood CE. 1949. The similarity paradox in human learning: a resolution. *Psychol Rev* **56:** 132–143. doi:10.1037/h0057488

Pastötter B, Bäuml KH. 2014. Retrieval practice enhances new learning: the forward effect of testing. *Front Psychol* **5:** 286. doi:10.3389/fpsyg.2014.00286

Pastötter B, Schicker S, Niedernhuber J, Bäuml KH. 2011. Retrieval during learning facilitates subsequent memory encoding. *J Exp Psychol Learn Mem Cogn* **37:** 287–297. doi:10.1037/a0021801

Postman L, Gray W. 1977. Maintenance of prior associations and proactive inhibition. *J Exp Psychol Hum Learn Mem* **3:** 255–263. doi:10.1037/0278-7393.3.3.255

Putnam AL, Sungkhasettee VW, Roediger HL III. 2017. When misinformation improves memory: the effects of recollecting change. *Psychol Sci* **28:** 36–46. doi:10.1177/0956797616672268

R Core Team. 2021. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Richter FR, Chanales AJ, Kuhl BA. 2016. Predicting the integration of overlapping memories by decoding mnemonic processing states during learning. *Neuroimage* **124:** 323–335. doi:10.1016/j.neuroimage.2015.08.051

Rindal EJ, DeFranco RM, Rich PR, Zaragoza MS. 2016. Does reactivating a witnessed memory increase its susceptibility to impairment by subsequent misinformation? *J Exp Psychol Learn Mem Cogn* **42:** 1544–1558. doi:10.1037/xlm0000265

Robbins D, Bray JF. 1974. Repetition effects and retroactive facilitation: immediate and delayed recall performance. *Bull Psychon Soc* **3:** 347–349. doi:10.3758/BF03333492

Roediger HL, Karpicke JD. 2006. Test-enhanced learning: taking memory tests improves long-term retention. *Psychol Sci* **17:** 249–255. doi:10.1111/j.1467-9280.2006.01693.x

Rowland CA. 2014. The effect of testing versus restudy on retention: a meta-analytic review of the testing effect. *Psychol Bull* **140:** 1432–1463. doi:10.1037/a0037559

Sahakyan L, Hendricks HE. 2012. Context change and retrieval difficulty in the list-before-last paradigm. *Mem Cognit* **40:** 844–860. doi:10.3758/s13421-012-0198-0

Schlichting ML, Preston AR. 2014. Memory reactivation during rest supports upcoming learning of related content. *Proc Natl Acad Sci* **111:** 15845–15850. doi:10.1073/pnas.1404396111

Schlichting ML, Preston AR. 2015. Memory integration: neural mechanisms and implications for behavior. *Curr Opin Behav Sci* **1:** 1–8. doi:10.1016/j.cobeha.2014.07.005

Scully ID, Hupbach A. 2020. Different reactivation procedures enable or prevent episodic memory updating. *Hippocampus* **30:** 806–814. doi:10.1002/hipo.23159

Simmons JP, Nelson LD, Simonsohn U. 2012. A 21-word solution. *Dialogue Can Philos Assoc* **26:** 4–7.

Smith SM, Vela E. 2001. Environmental context-dependent memory: a review and meta-analysis. *Psychon Bull Rev* **8:** 203–220. doi:10.3758/BF03196157

Szpunar KK, McDermott KB, Roediger HL III. 2008. Testing during study insulates against the buildup of proactive interference. *J Exp Psychol Learn Mem Cogn* **34:** 1392–1399. doi:10.1037/a0013082

Tousignant JP, Hall D, Loftus EF. 1986. Discrepancy detection and vulnerability to misleading postevent information. *Mem Cognit* **14:** 329–338. doi:10.3758/BF03202511

Tulving E, Watkins MJ. 1974. On negative transfer: effects of testing one list on the recall of another. *J Verb Learn Verb Behav* **13:** 181–193. doi:10.1016/S0022-5371(74)80043-5

Wahlheim CN. 2015. Testing can counteract proactive interference by integrating competing information. *Mem Cognit* **43:** 27–38. doi:10.3758/s13421-014-0455-5

Wahlheim CN, Jacoby LL. 2013. Remembering change: the critical role of recursive remindings in proactive effects of memory. *Mem Cognit* **41:** 1–15. doi:10.3758/s13421-012-0246-9

Wahlheim CN, Smith WG, Delaney PF. 2019. Reminders can enhance or impair episodic memory updating: a memory-for-change perspective. *Memory* **27:** 849–867. doi:10.1080/09658211.2019.1582677

Wahlheim CN, Garlitch SM, Kemp PL. 2021. Context differentiation and remindings in episodic memory updating. *Psychol Learn Motiv* **75:** 245–277. doi:10.1016/bs.plm.2021.06.001

Warren DE, Jones SH, Duff MC, Tranel D. 2014. False recall is reduced by damage to the ventromedial prefrontal cortex: implications for understanding the neural correlates of schematic memory. *J Neurosci* **34:** 7677–7682. doi:10.1523/JNEUROSCI.0119-14.2014

Wickham H, Averick M, Bryan J, Chang W, McGowan LDA, François R, Grolemund G, Hayes A, Henry L, Hester J, et al. 2019. Welcome to the Tidyverse. *J Open Source Softw* **4:** 1686. doi:10.21105/joss.01686

Yang C, Potts R, Shanks DR. 2018. Enhancing learning and retrieval of new information: a review of the forward testing effect. *NPJ Sci Learn* **3:** 8. doi:10.1038/s41539-018-0024-y

Received April 20, 2023; accepted in revised form July 10, 2023.

Interpolated retrieval retroactively increases recall and promotes cross-episode memory interdependence

Christopher N. Wahlheim, Sydney T. Smith, Sydney M. Garlitch, et al.

Learn. Mem. 2023, **30:**
Access the most recent version at doi:10.1101/lm.053782.123