Testing can counteract proactive interference by integrating competing information

By: Christopher N. Wahlheim

Wahlheim, C. N. (2015). Testing can counteract proactive interference by integrating competing information. *Memory & Cognition, 43(1),* 27-38.doi:10.3758/s13421-014-0455-5

Abstract:

Testing initially learned information before presenting new information has been shown to counteract the deleterious effects of proactive interference by segregating competing sources of information. The present experiments were conducted to demonstrate that testing can also have its effects in part by integrating competing information. Variations of classic A–B, A–D paired-associate learning paradigms were employed that included two lists of word pairs and a cued-recall test. Repeated pairs appeared in both lists (A–B, A–B), control pairs appeared in List 2 only (A–B, C–D), and changed pairs appeared with the same cue in both lists but with different responses (A–B, A–D). The critical manipulation was whether pairs were tested or restudied in an interpolated phase that occurred between Lists 1 and 2. On a final cued-recall test, participants recalled List 2 responses and then indicated when they recollected that responses had earlier changed between lists. The change recollection measure indexed the extent to which competing responses were integrated during List 2. Change was recollected more often for tested than for restudied pairs. Proactive facilitation was obtained in cued recall when change was recollected, whereas proactive interference was obtained when change was not recollected. These results provide evidence that testing counteracted proactive interference in part by making List 1 responses more accessible during List 2, thus promoting integration and increasing later recollection of change. These results have theoretical implications because they show that testing can counteract proactive interference by integrating or segregating competing information.

Keywords: Proactive interference | Testing effects | Change recollection | Reminding | Integration

Articles:

A large body of evidence has shown that by promoting the act of retrieval, testing can enhance memory in several ways (for reviews, see Roediger & Butler, 2011; Roediger & Karpicke, 2006a). One way that testing can enhance memory is by counteracting the deleterious effects of proactive interference. Proactive interference occurs when prior learning impairs memory for more recently learned information (for reviews, see Anderson & Neely, 1996;

Crowder, 1976; Postman & Underwood, 1973). The primary mechanism by which testing has been held to counteract proactive interference is by segregating competing sources of information (e.g., Szpunar, McDermott, & Roediger, 2008; Tulving & Watkins, 1974). Consistent with this view, testing has been shown to induce context change (Pastötter, Schicker, Niedernhuber, & Bäuml, 2011; see also Jang & Huber, 2008), reduce cue overload (Nunes & Weinstein, 2012), and increase test expectancy (Weinstein, Gilmore, Szpunar, & McDermott, 2014). Although segregating competing information has long been shown to be effective for counteracting proactive interference (e.g., Underwood & Ekstrand, 1967), testing might also have its effects in an opposite way. The purpose of the present study is to show that testing can counteract proactive interference by promoting the integration of competing information.

The effects of testing on proactive interference have most prominently been demonstrated in multiple-list learning paradigms (e.g., Darley & Murdock, 1971; Nunes & Weinstein, 2012; Pastötter et al., 2011; Szpunar et al., 2008; Weinstein et al., 2014). For example, in Szpunar et al. (2008), participants studied five lists of words that were sometimes related and sometimes unrelated across lists. The primary manipulation was the task that intervened between study lists. Following each study list, groups of participants were either given a recall test or they were not tested. Groups that were not tested were either given additional study opportunities or completed distractor tasks. Following study, all participants were given a recall test on the final list following a short filled delay and a cumulative recall test over all the lists following a longer delay. The critical comparison was performance on both of these recall tasks for tested and nontested groups. Evidence that testing counteracted proactive interference was shown on the test of the final list by more words being correctly recalled for tested than nontested groups, and more intrusions from prior lists for nontested than tested groups. Similar evidence was shown on the final cumulative recall test in that more words from the entire experiment were correctly recalled for tested than nontested groups. These results were interpreted as showing that testing improved source monitoring by segregating lists.

Another paradigm that has been used to demonstrate that testing can counteract proactive interference is the A–B, A–D paired-associate learning paradigm. For example, in Tulving and Watkins (1974), participants were instructed to study two lists of word pairs that contained left members (cues) that were the same in both lists paired with right members (responses) that changed between lists. The critical manipulation was whether or not immediate cued-recall tests followed Lists 1 and 2. In the final phase, participants completed a modified free recall (MMFR) test in which cues from the studied lists appeared with instructions to recall responses from both lists. Including an immediate test of List 1 resulted in List 2 responses being better recalled on both List 2 and MMFR tests than when a List 1 test was not included. However, the effect was only obtained when the entire list was tested in a between-subjects design. When only a subset of items was tested in a within-subjects design, proactive interference did not differ for tested and nontested items. These results were interpreted as showing that testing List 1 in its entirety was required to segregate it from the subsequent presentation of List 2, whereas list segregation could not be accomplished when only a portion of the List 1 items was tested.

The paradigms described above can produce two types of proactive interference effects: list-level and item-level. In multiple-list learning paradigms (e.g., Szpunar et al., 2008), proactive interference has list-level effects in that the presentation of prior lists impairs memory for subsequent lists, but effects on individual items cannot be directly assessed. In contrast, in A–B, A–D paradigms (e.g., Tulving & Watkins, 1974), proactive interference can have both list- and item-level effects. Item-level effects are often of primary interest in A–B, A–D paradigms, and they are shown when memory for List 2 pairs is worse for pairs with responses that changed from List 1 to List 2 than control pairs with only one response in List 2. Outside of the lab, multiple-list learning paradigms represent situations in which earlier blocks of learning create global impairment in memory for later blocks of learning, such as when a student shows impaired memory for information from the latter of back-to-back classes. In contrast, the A–B, A–D paradigm represents situations in which changes in specific associations impair memory for more recent associations, such as when one attempts to remember an acquaintance’s married name after having previously known the acquaintance by her maiden name.

The distinction between these types of proactive interference is important for understanding how testing can have its effects. In both cases, segregating sources of information could make the sources more distinguishable, thus decreasing proactive interference. However, segregation might not be the only optimal strategy for paired-associate learning. Effective segregation of overlapping associations (A–B, A–D) results in traces for each association being represented independently. Under these circumstances, retrieval of target associations requires precise reinstatement of context to avoid retrieving competing associations. However, if the reinstatement of context is impoverished, then retrieval of shared cues can have the undesirable consequence of eliciting unwanted responses. Thus, effective segregation requires creating sufficiently distinct contexts at encoding. An alternative strategy is to integrate competing associations by specifying their relationship to one another so that the associations will be represented together with their relationship in a unified trace. During encoding, this would involve considering both associations in the same context along with their relationship to one another. A broader reinstatement of context could be sufficient to provide access to the integrated trace, and relational information could be used distinguish associations postretrieval.

It is reasonable to hypothesize that interpolated testing promotes integration in A–B, A–D paradigms because retrieval practice of List 1 pairs makes them more accessible when changed pairs are studied in List 2. In this vein, Wahlheim and Jacoby (2013) recently examined the role of a retrieval-based mechanism (i.e., reminding) in the integration of competing responses in an A–B, A–D paradigm. In their study, participants were presented with two lists of word pairs with repeated pairs appearing in both lists (A–B, A–B), control pairs appearing only in the second list (A–B, C–D), and changed pairs appearing with the same cue in both lists and a different response in each list (A–B, A–D). After learning List 1, participants studied List 2 and were told to indicate when they noticed changed pairs, a form of reminding referred to as change detection. On a final cued-recall test of List 2, a remindings-report procedure was employed in which participants attempted to recall List 2 responses, and then reported whether another word came to mind prior to their response. Instances in which List 1 responses were reported as coming to mind prior to responses output were taken as evidence that the change detected in List 2 had been recollected at test. These instances were validated as reflecting integrated traces in that they occurred exclusively for items detected as changed in List 2. Recall of List 2 responses did not differ overall between A–B, A–D and A–B, C–D items, showing no evidence for proactive interference. This lack of difference resulted from performance on A–B, A–D items constituting a mixture of proactive facilitation and proactive interference that depended on whether change had been recollected after it had been detected. Proactive facilitation was obtained when change was detected and recollected in that recall performance was greater for A–B, A–D than A–B, C–D items. In contrast, proactive interference was obtained when change was detected but not recollected in that recall performance was lower for A–B, A–D than A–B, C–D items.

To explain these results, Wahlheim and Jacoby (2013) forwarded a *memory-for-change* account that held that when change was detected in List 2, the retrieval of a List 1 response facilitated the formation of an integrated trace that included responses from each list along with the retrieval event that fostered their integration (for similar accounts of the role of study-phase retrieval in memory for temporal order, see Hintzman, 2010; Tzeng & Cotton, 1980; Winograd & Soloway, 1985). Change recollection enhanced memory performance because integrated traces containing both responses and their relative study order could be retrieved. In contrast, when change was detected but not recollected, memory performance was impaired because earlier retrieval of the List 1 competitor made it more accessible, and thus more likely to compete with the target response from List 2. In the parlance of dual-process theory (e.g., Jacoby, 1991), change recollection opposed the enhanced strength of List 1 competitors resulting from change detection, but in its absence, the greater strength of List 1 was the primary basis for responding.

Following this, Jacoby and Wahlheim (2014) provided a more direct demonstration of the role of change detection in the integration of responses in an A–B, A–D paradigm. To do this, they adopted a variant of the looking back procedure (Jacoby, 1974) to manipulate the extent to which participants looked for change that occurred between Lists 1 and 2. Their paradigm was similar to that of Wahlheim and Jacoby (2013), except that pairs that repeated across lists (A–B, A–B) were replaced by A–B, A–D pairs that changed within List 2, and the remindings-report procedure at test was replaced by a yes/no change recollection measure. During study of List 2, groups of participants were instructed either to look for pairs that changed from any point earlier in the experiment or to only look for pairs that changed from earlier in List 2. Of primary interest was how List 2 recall and change recollection for A–B, A–D items that changed between lists (between-list A–B, A–D items), differed as a function of looking-back instructions (for a similar procedure used to examine spaced repetition effects, see Wahlheim, Maddox, & Jacoby, 2014). For the group that looked for change within List 2, the results revealed no difference in List 2 recall between A–B, C–D and between-list A–B, A–D items. In contrast, for the group who looked for change in both lists, List 2 recall was greater for between-list A–B, A–D items than for A–B, C–D items, showing proactive facilitation in overall performance. Change recollection for between-list A–B, A–D items was greater for the group who looked for change in both lists, showing that more integrated traces were established by encouraging retrieval of List 1 A–B pairs when studying List 2 A–D pairs. Finally, replicating Wahlheim and Jacoby (2013) proactive facilitation was found for A–B, A–D pairs when change was detected and recollected, whereas proactive interference was found when change was detected but not recollected. These results showed that the benefits of change recollection on recall were due to trace integration fostered by change detection.

The findings of Jacoby and Wahlheim (2014) have important theoretical implications, in that a segregation account would predict that encouraging retrieval of competing responses from a previous list would further impair memory by increasing proactive interference. In contrast, just the opposite was found: Encouraging retrieval of competing information resulted in proactive facilitation in overall recall performance. Testing List 1 competitors prior to List 2 could serve a function similar to that of encouraging retrieval of List 1 during the presentation of List 2. It has recently been suggested that completing a test trial increases the likelihood of study-phase retrieval on a subsequent study trial (S. M. Nelson, Arnold, Gilmore, & McDermott, 2013). Taken with the increased accessibility of List 1 pairs produced by retrieval practice, testing List 1 should increase the contact made between A–B and corresponding A–D pairs during List 2 study.

The present experiments build on the previous work by Wahlheim and Jacoby by exploring the effects of interpolated testing on change recollection and its effects on List 2 recall. Specifically, an A–B, A–D paradigm was employed here to explore the possibility that testing memory for List 1 pairs prior to presenting List 2 can counteract proactive interference in a manner similar to directing retrieval to List 1 while studying List 2 (Jacoby & Wahlheim, 2014). The three item types used by Wahlheim and Jacoby (2013) were employed here: pairs repeated across lists (A–B, A–B); pairs appeared only in List 2 (A–B, C–D); and pairs with the same cue but responses that changed from List 1 to List 2 (A–B, A–D). The critical manipulation was the task interpolated between Lists 1 and 2. The interpolated task included (a) tested items that were cues for the recall of List 1 responses and (b) restudied items that were repeated presentations of List 1 items. Importantly, the interpolated task was manipulated within subjects, such that only half of List 1 was tested. This was done because a segregation account holds that the entire list must be tested in order to counteract proactive interference (e.g., Tulving & Watkins, 1974). Thus, if testing counteracts proactive interference when only a portion of the list is tested, its effects cannot be explained by a segregation account. The effects of interpolated testing were compared with those of interpolated study to ensure that testing benefits would not simply be due to increased exposure during the interpolated task. A change detection measure was not employed in List 2 because it would encourage the retrieval of List 1 responses, which could diminish differences between the interpolated task conditions. A yes–no change recollection measure was included on the final cued-recall test.

Proactive interference was expected for restudied but not for tested items, for reasons described above. Testing effects were expected to be due to the promotion of integrated traces, as shown by greater change recollection for tested than for restudied items. Finally, consistent with the *memory-for-change* account, proactive facilitation was expected when change was recollected, and proactive interference was expected when change was not recollected, consistent with several earlier studies (Jacoby & Wahlheim, 2014; Jacoby, Wahlheim, & Yonelinas, 2013; Putnam, Wahlheim, & Jacoby, 2014; Wahlheim, 2014; Wahlheim & Jacoby, 2013).

Experiment 1

Experiment 1 was designed to demonstrate the positive effects of List 1 testing on change recollection and proactive interference in an A–B, A–D paradigm. The design was similar to that of Wahlheim and Jacoby (2013), but different materials were used. Wahlheim and Jacoby had used three-word sets that each included one cue (e.g., *knee*) and two responses (e.g., *bone*, *bend*). Some of the cues and responses were semantically related, resulting in low average associations among the terms, and all responses were orthographically related, because they were originally intended to complete the same word fragment (e.g., *b_n_*). In contrast, the three-word sets used here included cues that were semantically related to the responses, but the responses were neither semantically nor orthographically related to one another (e.g., *pearl–harbor*, *pearl–jewelry*). Changing the materials in this way served two purposes. First, pilot studies had shown that these materials produced proactive interference in overall performance when List 1 was not tested, which was necessary for determining whether List 1 testing would eliminated proactive interference. Second, mediation accounts of failures to find interference effects (e.g., Barnes & Underwood, 1959) typically attributed those failures to spreading activation between related responses increasing memory performance. Here, a mediation account could not explain variations in proactive effects of memory that depend on change recollection, because the response terms were unrelated (see also Wahlheim, 2014). Finally, another difference was that change recollection judgments were made on a continuous scale. This was done to determine whether change recollection certainty would differ on the basis of the type of interpolated task.

Method

Participants

A group of 48 students from Washington University participated in exchange for $10/h or partial course credit. Participants were tested individually.

Design and materials

A 3 (Item Type: A–B, A–B vs. A–B, C–D vs. A–B, A–D) × 2 (Interpolated Task: test vs. restudy) within-subjects design was used. The materials consisted of 90 three-word sets, each including a cue word (e.g., *pearl*) and two responses that were associated with the cue but not with each other (e.g., *harbor*, *jewelry*). The average forward and backward associative strengths between cues and responses were low, on average (forward, M = .04, SD = .02, range = .01–.09; backward, M = .02, SD = .03, range = .00–.18), as indexed by D. L. Nelson, McEvoy, and Schreiber (1998). The materials were divided into six groups of 15 sets and rotated through conditions, resulting in six experimental formats. The groups served equally often across participants.

The experiment consisted of four phases: List 1, Interpolated Task, List 2, and List 2 Test (see Table 1 for a schematic of the design). List 1 consisted of 60 word pairs that were divided evenly between two groups that later became A–B, A–B or A–B, A–D items (30 each). Those groups were evenly divided into subgroups that were later restudied or tested in the interpolated task (15 each). The interpolated task consisted of the 60 pairs presented in List 1 divided into groups as just described. List 2 consisted of 90 pairs, divided evenly across the three items types (30 each), with half of the A–B, A–B and A–B, A–D items earlier having been restudied and the other half having been tested (15 each). The 30 A–B, C–D items were not subjected to the interpolated task manipulation. A–B, A–B items were repetitions of pairs that had appeared in List 1 (e.g., *baby–cute*, *baby–cute*); A–B, C–D items appeared for the first and only time in List 2 (e.g., *soup–bowl*); and A–B, A–D items appeared with cues from List 1 paired with changed responses in List 2 (e.g., *pearl–harbor*, *pearl–jewelry*). For A–B, A–D items, the designation of responses to lists remained constant across formats.

Table 1. General design schematic for Experiments 1 and 2

Item Type	List 1 (60 Items)	Interpolated Task (60 Items)	List 2 (90 Items)	List 2 Test (90 Items)
A–B, A–B	30 A–B items; repeated in List 2 (A–B)	15 test (A–?)	30 A–B items	15 test (A–?)
		15 restudy (A–B)		15 restudy (A–?)
A–B, A–D	30 A–B items; changed in List 2 (A–D)	15 test (A–?)	30 A–D items	15 test (A–?)
		15 restudy (A–B)		15 restudy (A–?)
A–B, C–D			30 C–D items	30 (C–?)

Procedure

In List 1, word pairs appeared individually in random order for 4 s each followed by a 1-s interstimulus interval (ISI). Participants were told to read the pairs for an upcoming test. In the interpolated task, restudied pairs appeared just as in List 1, whereas tested pairs appeared as cues paired with question marks (e.g., *pearl–?*). Pairs appeared individually in random order for 4 s each, followed by a 1-s ISI. Participants were told to read restudied pairs aloud and to study them, and to retrieve the List 1 responses for tested pairs before the 4 s expired. Responses were made aloud and written down by an experimenter. In List 2, the word pairs appeared individually in random order for 4 s each, followed by a 1-s ISI. Participants were told to read them aloud and study them for an upcoming memory test. At test, cues paired with question marks again appeared individually in random order, and participants were told to recall the responses presented in List 2 and to type them into the computer. Participants were encouraged to make a response for every item, guessing if necessary. After attempting to recall the List 2 response, participants indicated how certain they were that a pair had earlier changed from List 1 to List 2. A sliding scale appeared beneath each item with a slider directly in the middle. Participants were told to move the slider left to indicate that the item had not changed, to the right to indicate that the item had changed, and to leave the slider in the middle to indicate uncertainty regarding change. No values were assigned to the scale from the participants' perspective, but participants were told that more extreme movements to the left and right indicated greater certainty. Participants were told to make their ratings precisely and to use the full range of the scale.

Results and discussion

The criterion for statistical significance was set at alpha = .05. Main effects that were qualified by significant interactions are not reported. Variations in degrees of freedom are due to the exclusion of participants without at least one observation in every cell in conditional analyses.

Cued recall

Table 2 shows that testing List 1 items in the interpolated phase counteracted proactive interference. Cued recall of List 2 responses on the final test did not differ between tested A–B, A–D items and A–B, C–D items, $t(47) = 0.35$, $p = .73$, but recall was lower for restudied A–B, A–D items than for A–B, C–D items, $t(47) = –3.76$, $p < .001$. List 2 recall of A–B, A–D items was also lower in the restudy than in the test condition, $t(47) = –2.67$, $p = .01$. More evidence that testing counteracted proactive interference was shown by List 1 intrusions for A–B, A–D items being substantially lower on the final recall test for tested than for restudied items (.20 vs. .34), $t(47) = –5.67$, $p < .001$. In addition, repeating items across lists increased performance for both interpolated conditions. Final recall was higher for A–B, A–B than for A–B, C–D items in both interpolated task conditions, smallest $t(47) = 14.26$, $p < .001$. Final recall was also higher for restudied than for tested A–B, A–B items, $t(47) = 2.68$, $p = .01$, which is consistent with earlier studies that had shown a recall advantage for restudying over testing when the final test occurred after a short delay (e.g., Roediger & Karpicke, 2006b). The advantage for restudied over tested A–B, A–B items was due to testing only benefitting items that were correctly recalled in the interpolated phase. Final recall of A–B, A–B items in the tested condition was dramatically greater for items correctly recalled in the interpolated phase than for those that were incorrectly recalled (.97 vs. .63), $t(47) = 9.41$, $p < .001$. Finally, no difference in List 1 recall emerged for pairs in the interpolated task that eventually became A–B, A–B or A–B, A–D items (.49 vs. .49), $t(47) = –0.10$, $p = .92$. Most critical to the issue of testing effects on proactive interference, these results show that testing counteracted proactive interference even when only a portion of the list was tested, which cannot be explained by a segregation account (cf. Tulving & Watkins, 1974). The presence of restudied items in the interpolated task discouraged segregation by increasing the likelihood that contact would be made between Lists 1 and 2.

Table 2. Cued recall of List 2 responses as a function of item type and interpolated task: Experiments 1 and 2

		Item Type		
Experiment	**Interpolated Task**	**A–B, A–B**	**A–B, C–D**	**A–B, A–D**
Experiment 1	Test	.77 (.02)	.48 (.02)	.47 (.03)
	Restudy	.81 (.02)	.48 (.02)	.40 (.03)
Experiment 2	Test	.89 (.02)	.57 (.03)	.49 (.04)
	Restudy	.87 (.02)	.57 (.03)	.47 (.04)

The A–B, C–D items did not differ for the interpolated task conditions. Consequently, List 2 recall performance for those items is displayed twice for comparison with the other items. Standard errors of the means are presented in parentheses.

Change recollection

Change recollection rates were compared for the interpolated task conditions to determine whether testing eliminated proactive interference by enhancing the integration of competing responses. The change recollection measure recorded responses numerically on a scale from .00 to 1.00. Responses made to the left of the midpoint of the scale, indicating that change was not recollected, ranged from .00 to .49. Responses made to the right of the midpoint, indicating that change was recollected, ranged from .51 to 1.00. Responses made on the midpoint, indicating uncertainty about change recollection, were recorded as .50. Change recollection was indexed as the probability of responses occurring to the right of the midpoint (>.50), and failure to recollect change was indexed as the probability of responses occurring to the left of the midpoint (<.50), both regardless of the values.

The change recollection results provide evidence that testing List 1 in the interpolated phase promoted the integration of competing responses for A–B, A–D items. Table 3 shows that change was recollected more often for tested than for restudied A–B, A–D items. Comparisons of change recollection rates for A–B, A–D items took into account false alarms on both other item types. The first analysis of change recollection rates included A–B, A–B false alarms. The advantage in change recollection rates for tested as compared to restudied items was greater for A–B, A–D than for A–B, A–B items, $F(1, 47) = 4.51$, $p = .04$, $\eta_p^2 = .09$. The second analysis of change recollection rates included A–B, C–D false alarms. Given that A–B, C–D items were not subjected to the interpolated task manipulation, change recollection rates for both types of A–B, A–D items were compared individually with the overall false alarm rate for A–B, C–D items. Change recollection rates were higher for A–B, A–D items than for A–B, C–D items, and those rates were higher for tested than for restudied A–B, A–D items, smallest $t(47) = 5.20$, $p < .001$. The higher rates of change recollection for tested A–B, A–D items depended on the accuracy of List 1 recall in the interpolated phase. Change recollection was greater following correct than following incorrect recall of List 1 responses in the interpolated phase (.70 vs. .49), $t(47) = 3.93$, $p < .001$, and this presumably reflects change being detected more often during List 2 when List 1 responses had been recalled. Change recollection did not differ between tested items that were not recalled in the interpolated phase and restudied items (.49 vs. .45), $t(47) = 1.03$, $p = .31$. Finally, change recollection certainty was compared between tested and restudied A–B, A–D items. The average certainty was higher for tested than for restudied items (.63 vs. .51), $t(47) = 4.85$, $p < .001$, but this merely reflected change being recollected more often for tested items. When change was recollected, the average certainty did not differ between the test and restudy conditions (.90 vs. .91), $t(46) = 1.19$, $p = .24$.

Table 3. Probabilities of change recollection as a function of item type and interpolated task: Experiments 1 and 2

		Item Type		
Experiment	**Interpolated Task**	**A–B, A–B**	**A–B, C–D**	**A–B, A–D**
Experiment 1	Test	.18 (.02)	.15 (.02)	.60 (.03)
	Restudy	.10 (.02)	.15 (.02)	.45 (.03)
Experiment 2	Test	.06 (.01)	.05 (.01)	.68 (.04)
	Restudy	.06 (.02)	.05 (.01)	.50 (.04)

The change recollection probabilities reflect all levels of change recollection certainty. The A–B, C–D items did not differ for the interpolated task conditions and are presented twice for comparison with the other items. Standard errors of the means are presented in parentheses.

Cued recall conditionalized on change recollection

List 2 recall conditionalized on change recollection showed that testing eliminated proactive interference by increasing change recollection, which produced proactive facilitation in the recall of A–B, A–D items. Replicating earlier studies, Table 4 shows that overall List 2 recall performance for A–B, A–D items reflected a mixture of proactive facilitation when change was recollected and proactive interference when change was not recollected. Cued recall was greater for A–B, A–D items when change was recollected than for A–B, C–D items in both interpolated task conditions, smallest $t(45) = 2.75$, $p = .009$, and cued recall for those A–B, A–D items did not differ between interpolated task conditions, $t(45) = 0.02$, $p = .98$. In contrast, cued recall was lower for A–B, A–D items when change was not recollected than for A–B, C–D items, smallest $t(45) = –5.07$, $p < .001$, and cued recall for those A–B, A–D items did not differ between interpolated task conditions, $t(45) = 1.59$, $p = .12$. Importantly, the magnitudes of proactive facilitation and proactive interference effects did not differ for tested and restudied items, showing that the recall advantage for tested A–B, A–D items could be completely accounted for by the increase in change recollection.

Table 4. Cued recall of List 2 responses conditionalized on change recollection, as a function of item type and interpolated task: Experiments 1 and 2

		Item Type		
Experiment	**Interpolated Task**	**A–B, A–D, Change Recollected**	**A–B, C–D**	**A–B, A–D, Change Not Recollected**
Experiment 1	Test	.58 (.04)	.48 (.02)	.28 (.04)
	Restudy	.58 (.04)	.48 (.02)	.22 (.03)
Experiment 2	Test	.65 (.04)	.57 (.03)	.09 (.03)
	Restudy	.61 (.04)	.57 (.03)	.28 (.04)

The A–B, C–D items did not differ for the interpolated task conditions. Consequently, List 2 recall for those items is displayed twice for comparison with the other items. Standard errors of the means are presented in parentheses.

Finally, the memory-for-change account of proactive effects of memory was supported by the results from List 1 intrusions rates on the final recall test. As was described in the introduction, the account holds that change recollection serves to oppose the accessibility of List 1 competitors on a final test of List 2 responses. Thus, the account predicts that when List 1 responses are correctly recalled for items tested in the interpolated phase, List 1 intrusions on the final test should be far less likely when change is recollected, because change recollection preserves the temporal order of responses. Consistent with this account, when List 1 responses were correctly recalled in the interpolated phase, the probability of List 1 intrusions on the final test for A–B, A–D items was dramatically lower when change was recollected than when it was not (.20 vs. .79), $t(35) = –7.49$, $p < .001$. Together, these results show that testing eliminated proactive interference in part by enhancing the integration of competing responses and the later recollection of change that opposed the accessibility of competitors from List 1.

Experiment 2

Experiment 2 was conducted to replicate the effects of interpolated testing on change recollection and the effects of change recollection on List 2 recall performance. Experiment 2 was also conducted to explore the boundary conditions of these effects by including feedback following testing in the interpolated task. The positive effects of testing on later memory performance can be enhanced by providing feedback, especially for information that cannot initially be recalled (e.g., Butler, Karpicke, & Roediger, 2008; Butler & Roediger, 2008; Karpicke & Roediger, 2007; McDaniel & Fisher, 1991). In Experiment 1, change recollection for tested A–B, A–D items was greater for items that were correctly recalled during the interpolated phase than for those that were not. Consequently, the provision of feedback could enhance participants' ability to detect change for items that were not recalled initially, leading to greater change recollection and enhanced overall recall of List 2 responses on the final test. However, the increased accessibility of List 1 responses resulting from feedback does not guarantee a proportional increase in change recollection at test. An alternative possibility is that feedback could increase the accessibility of the List 1 responses, which would enhance the detection of change, but without a corresponding increase in the recollection of change the enhanced accessibility of competitors would hurt memory performance more when change was not recollected. A similar result was shown by Wahlheim (2014), in that increasing the number of List 1 presentations increased change recollection but also had the offsetting effect of producing greater proactive interference when change was not recollected.

Another change made in Experiment 2 was that the change recollection measure was no longer continuous. This was done to simplify analyses, because the continuous measure in Experiment 1 did not reveal any differences between interpolated task conditions. Experiment 2 included a dichotomous yes–no self-report measure of change recollection that followed the List 2 recall attempts. When participants reported recollecting change, they were subsequently asked to recall the List 1 response. If testing with feedback increases the accessibility of List 1 items beyond those that were restudied, then List 1 recall following change recollection should be greater for tested than for restudied items.

Method

Participants

A group of 36 students from Washington University participated in exchange for $10/h or partial course credit. Participants were tested individually.

Design, materials, and procedure

The design, materials, and procedure were identical to those of Experiment 1, with the following exceptions. The presentation duration for items presented in the interpolated phase was 6 s. For tested items, cues appeared with question marks for 4 s, during which time participants were told to recall the response from List 1. The correct response then appeared below for an additional 2 s. Participants were told to respond to every item. This was done to encourage retrieval attempts for all items and to prevent the strategy of waiting for feedback to appear. At test, change recollection was assessed using a dichotomous yes–no self-report measure. After participants had typed in their List 2 recall attempts, the message "Did the right word change from List 1 to List 2?" appeared above boxes labeled "Yes" and "No." Participants clicked on the boxes to enter their responses. When participants indicated that they recollected change, they were then asked to type in the response that had appeared in List 1.

Results and discussion

Cued recall

In contrast to Experiment 1, the bottom rows of Table 2 show that proactive interference was obtained in overall recall of List 2 responses for both types of A–B, A–D items. List 2 recall was lower for both tested and restudied A–B, A–D items than for A–B, C–D items, smallest $t(35) = -3.25$, $p = .003$, and did not differ between A–B, A–D item types, $t(35) = 0.53$, $p = .60$. Only a marginally significant difference in List 1 intrusions emerged on the final recall test between tested and restudied A–B, A–D items (.25 vs. .29), $t(35) = 1.85$, $p = .07$. Consistent with Experiment 1, repeating items across lists increased performance for both interpolated conditions. Final recall was higher for A–B, A–B than for A–B, C–D items in both interpolated task conditions, smallest $t(35) = 12.99$, $p < .001$. However, unlike in Experiment 1, final recall for A–B, A–B items did not differ between interpolated task conditions, $t(35) = 1.25$, $p = .22$, and trended in the opposite direction, showing a numeric advantage for tested items, even though the final test was given immediately following List 2 study (cf. Roediger & Karpicke, 2006b). This was explained by final recall performance on tested A–B, A–B items showing a smaller difference between items that had been correctly and incorrectly recalled in the interpolated phase (.96 vs. .85), $t(35) = 3.55$, $p = .001$, relative to Experiment 1. Feedback in the interpolated phase increased the final recall of A–B, A–B items incorrectly recalled in the interpolated phase to a relatively greater extent than the recall of items correctly recalled in the interpolated phase, because performance on the latter condition was at ceiling. Finally, as in Experiment 1, no difference in List 1 recall was apparent in the interpolated phase for pairs that eventually became A–B, A–B or A–B, A–D items (.53 vs. .51), $t(35) = 0.60$, $p = .56$.

These results show that testing did not eliminate proactive interference in overall performance as in Experiment 1. However, despite the similarity in performance between the interpolated conditions for A–B, A–D items here in Experiment 2, it is likely that the overall levels of List 2 recall were accomplished in different ways. Testing might have counteracted proactive interference to a greater extent than restudy, but the inclusion of feedback could potentially have resulted in greater offsetting effects of proactive interference when change was not recollected, because of the increased accessibility of List 1 responses.

Change recollection

Change recollection rates were compared for the interpolated task conditions, as in Experiment 1. The change recollection results again provide evidence that testing List 1 promoted the integration of responses. The bottom rows of Table 3 show that change was recollected more often for tested than for restudied A–B, A–D items. The analysis of change recollection including A–B, A–B false alarms showed that the advantage in change recollection for tested as compared to restudied items was greater for A–B, A–D than for A–B, A–B items, $F(1, 35) = 41.77$, $p < .001$, $\eta_p^2 = .54$. In addition, the analysis of change recollection including A–B, C–D false alarms showed that change recollection was higher for A–B, A–D items than for A–B, C–D items, and that those rates were higher for tested than for restudied A–B, A–D items, smallest $t(35) = 6.61$, $p < .001$. In contrast to Experiment 1, change recollection for tested A–B, A–D items did not depend on correct recall of the List 1 responses in the interpolated phase, because feedback was provided. Change recollection did not differ for A–B, A–D items that were correctly or incorrectly recalled in the interpolated phase (.70 vs. .66), $t(35) = 0.94$, $p = .35$ showing that, as compared to Experiment 1, feedback selectively increased change recollection for A–B, A–D items that were incorrectly recalled in the interpolated phase.

Cued recall conditionalized on change recollection

As in Experiment 1, cued recall conditionalized on change recollection showed that testing counteracted proactive interference, despite the finding that proactive interference was obtained in overall recall performance. The bottom rows of Table 4 show that when change was recollected, proactive facilitation was obtained in recall of tested A–B, A–D items, but neither proactive facilitation nor proactive interference was obtained for restudied A–B, A–D items. Cued recall was greater for tested A–B, A–D items for which change was recollected than for A–B, C–D items, $t(33) = 2.49$, $p = .02$, and cued recall for restudied A–B, A–D items for which change was recollected did not differ from that for A–B, C–D items, $t(33) = 0.84$, $p = .41$; however, cued recall did not differ between tested and restudied A–B, A–D items when change was recollected, $t(35) = 0.64$, $p = .53$. In contrast, when change was not recollected, proactive interference was obtained in recall of both types of A–B, A–D items, with the magnitude of proactive interference being greater in the test than in the restudy condition. Cued recall was lower for both types of A–B, A–D items when change was not recollected than for A–B, C–D items, smallest $t(33) = –6.61$, $p < .001$. Cued recall was also lower for tested than for restudied A–B, A–D items when change was not recollected, $t(33) = –3.22$, $p = .003$. These results show that testing did not eliminate proactive interference in overall recall performance as in Experiment 1, because increasing the accessibility of List 1 competitors with feedback increased proactive interference when change was not recollected more than it increased proactive facilitation when change was recollected. Evidence that testing with feedback enhanced the accessibility of List 1 responses beyond restudy was shown by List 1 intrusions on A–B, A–D items on the final recall test being greater in the test than in the restudy condition when change was not recollected (.74 vs. .53), $t(33) = 3.20$, $p = .003$, and by List 1 responses being recalled more often following change recollection on A–B, A–D items in the test than in the restudy condition (.87 vs. .74), $t(35) = 4.01$, $p < .001$.

Finally, the memory-for-change account of proactive effects of memory was again supported by the List 1 intrusion rates. When the List 1 responses were correctly recalled in the interpolated phase, the probability of List 1 intrusions on the final test for A–B, A–D items was dramatically lower when change was recollected than when it was not (.09 vs. .87), $t(30) = –10.77$, $p < .001$. Together, these results again show that testing counteracted proactive interference in part by enhancing the integration of competing responses and the later recollection of change that opposed the accessibility of competitors.

General discussion

The present experiments demonstrated that testing counteracted proactive interference in part by enhancing the integration of competing responses. Response integration was indexed by change recollection, which presumably reflected memory for traces that were earlier integrated when change was detected during List 2 study. Change recollection occurred more often for A–B, A–D items that were tested in the interpolated phase than for those that were restudied. The provision of feedback in the interpolated phase tended to increase change recollection, but also had the negative effect of enhancing the strength of competitors and consequent proactive interference when change was not recollected, consistent with the effects of increasing List 1 presentations (Wahlheim, 2014). Together, these results suggest that one role of testing is to promote the integration of competing information, but the effects of testing on overall recall performance depend on the extent to which the detection of change can be later recollected.

The results from Experiment 1 differed from those of Tulving and Watkins (1974), in that they had found that interpolated testing did not counteract proactive interference when it was manipulated within subjects. The exact reason for this discrepancy is unclear, due to the variety of differences between studies. However, two primary differences provide a potential explanation. Tulving and Watkins used sets of words that were neither semantically nor orthographically related, and they only tested a quarter of the A–B pairs that corresponded to A–D pairs in a subsequent list. These conditions discouraged contact between lists and could have resulted in tested pairs behaving similarly to nontested pairs, because items were unlikely to be integrated. Furthermore, even if some items were integrated, offsetting effects of proactive interference in the absence of change recollection could have obscured benefits in overall recall performance from presenting. In contrast, the present study employed materials in which the cues were semantically related to the responses, but the responses were not related to each other. These materials had earlier produced change recollection rates comparable to those from materials for which all terms were semantically related, and both types of items produced change recollection rates that were substantially higher than the rates produced by unrelated sets (Wahlheim, 2014). In addition, half of the List 1 items were tested in the present study, which did more to encourage detection of change in List 2 than did testing only a quarter of the items. Together, these features made the conditions in the present study more conducive to integration, which likely contributed to the elimination of proactive interference effects in overall recall of List 2 responses for the subset of items tested in the interpolated phase following List 1.

The notion that integrative processes can counteract proactive interference is consistent with findings from several studies (e.g., Jacoby, 1974; Jacoby & Wahlheim, 2013, 2014; Jacoby et al., 2013; Putnam et al., 2014; Wahlheim, 2014; Wahlheim & Jacoby, 2013), and more generally with findings showing that integration counteracts interference produced by retrieval-induced forgetting (Anderson & McCulloch, 1999) and fan effects (Radvansky & Zacks, 1991). More relevant to the present study, several studies in the testing-effect literature have also shown indirect evidence that testing promotes integration in the form of test-enhanced memory for both studied materials and related but unstudied materials (for a review, see Carpenter, 2012). For example, testing has been shown to enhance the recall of unstudied prose material related to studied material (Chan, McDermott, & Roediger, 2006), the classification of novel exemplars of earlier-studied natural categories (Jacoby, Wahlheim, & Coane, 2010), knowledge-based inferences (Butler, 2010), and transfer of rules to novel materials within the same knowledge domain (Kang, McDaniel, & Pashler, 2011). More direct evidence that testing promotes integration was also shown by Wissman, Rawson, and Pyc (2011), in that testing in the interim between passages of connected discourse improved the comprehension of material beyond other interpolated tasks.

Although many studies have provided evidence that testing promotes integration, there is no general consensus on the mechanisms that underlie these effects. An important contribution of the present article is that it provides a potential account of how testing could promote integration. As was described in the introduction, the memory-for-change account holds that detecting change results from the retrieval of earlier-learned information during the learning of new information. The information from both occasions then becomes integrated into a unitary representation. Later recollection of change provides access to the integrated representation that preserves the temporal relationship of information from the separate occasions. Competing sources of information can be distinguished using relational information within an integrated trace, but the representation of an integrated trace differs from those formed via segregation. Trace integration produces a unitary representation, whereas segregation produces distinct traces that are represented independently. Importantly, testing appears to increase the frequency of study-phase retrievals and later recollection of integrated representations. Although the memory-for-change account of testing effects on proactive interference in a paired-associate learning paradigm cannot fully explain other related phenomena showing integrative effects of testing, the account does point to a study-phase retrieval mechanism being critical for the integration of information. Moreover, the generality of a study-phase retrieval account of integrative effects of testing has been shown in studies demonstrating the positive effects of looking back for both paired associates (Jacoby & Wahlheim, 2014) and individual verbal category exemplars (Jacoby & Wahlheim, 2013). Also related, the detection of contradictions in text produces similar memorial benefits (Otero & Kintsch, 1992). It is possible that testing in situations using more complex materials could foster integrated representations by encouraging study-phase retrieval, but more research will be needed to determine the exact nature of those representations.

Despite the evidence for the role of an integrative mechanism in the effects of testing on proactive interference shown here, it is likely that other mechanisms also play important roles in these effects. The integrative mechanism forwarded here is likely to play a role in many situations, but the representations posited here might only be effectively formed when the number of competing units is manageable. In contrast, a segregation mechanism might play a more prominent role in situations that involve a greater number of competing responses. For example, Szpunar et al. (2008) showed that interpolating tests among five 18-word study lists improved free recall on a test of the final list and decreased intrusions as compared to when interpolated study or distractor tasks were employed. These effects were obtained both when several exemplars from various categories were distributed across lists and when unrelated words populated the lists. An integration mechanism of the sort proposed in the present study could not easily explain the effects on final recall performance, because the representations would have to preserve relational information across lists of items instead of between two responses paired with the same cue. In addition, study-phase retrievals that preserve relative order are unlikely to occur when the materials are unrelated words. There is also evidence that other processes contribute to the segregation of lists in the Szpunar et al. paradigm. For example, Pastötter et al. (2011) used an electroencephalograph measure to infer that attention was greater during study trials when retrieval tasks rather than study or distractor tasks were interpolated between lists, presumably due to shifts of context. Also, Nunes and Weinstein (2012) showed that testing increased correct recall of categorically related items without increasing false recall of related but unstudied lures, presumably by reducing cue overload. Finally, Weinstein, Gilmore, Szpunar, and McDermott (2014) showed that test expectancy on study trials was greater when tests rather than distractor tasks were interpolated, and this presumably facilitated subsequent encoding and later segregation of lists. These findings show that a comprehensive account of testing effects on proactive interference will require consideration of a variety of mechanisms and their interactions with learning conditions.

It is likely that the conditions in which testing is employed will play an important role in determining the primary mechanism that is responsible for its effects. In addition to situational influences, differences in learner-initiated goals might also influence the primary mechanism of testing effects. As a concrete example, consider students in an educational context. Testing with the goal of integration can be employed when students quiz themselves to prepare for exams or when students take exams in courses in which the concepts build throughout the semester. In contrast, testing with the goal of segregation can be employed when students need to offload information to increase the availability of transient neural activity required for directing attention to and sustaining attention on a different topic. The modulation of these uses for testing is likely to reflect individual differences in attentional control and the use of situation-appropriate learning strategies. It is possible that students who perform poorly due to deficits in attentional control might suffer when segregation is required because goal-directed attempts to forget previously retrieved information ironically make that information more accessible (cf. Jacoby, 1999; Wegner, 1994). This could result in students making distant connections among concepts following retrieval attempts, but would also lead to greater confusion resulting from interference among unrelated ideas. Examination of the interactions among attentional control, testing effects, proactive effects of memory, and generalization of knowledge would be very informative.

In conclusion, the present study provides evidence that testing can promote the integration of competing information and increase later recollection of integrated traces. This finding adds to the testing-effect literature by showing that testing can counteract proactive interference by integrating competing information. The function of testing likely depends on the situation in which it is used, the capability of individuals to control their attention, and the strategy motivating the use of testing. Future research should explore which variables influence the mechanism that is primarily responsible for testing effects on proactive interference, both for theoretical purposes and to optimize the use of testing to enhance student learning.

Notes

Author Note

This research was supported by a Collaborative Activity Award from the James S. McDonnell Foundation awarded to Larry Jacoby, and by Grant No. 5T32AG000030-38 from the National Institute on Aging. Thanks to Larry Jacoby for his substantial contributions to the ideas presented here, Ashley Bartels for her assistance with manuscript preparation and data collection, and Heather Bartels and Kara Golebiowski for their assistance with data collection.

References

1. Anderson, M. C., & McCulloch, K. C. (1999). Integration as a general boundary condition on retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 608–629. doi: 10.1037/0278-7393.25.3.608
2. Anderson, M. C., & Neely, J. H. (1996). Interference and inhibition in memory retrieval. In E. L. Bjork & R. A. Bjork (Eds.), *Memory (Handbook of Perception and Cognition)* (2nd ed., pp. 237–313). San Diego, CA: Academic Press.
3. Barnes, J. M., & Underwood, B. J. (1959). "Fate" of first-list associations in transfer theory. *Journal of Experimental Psychology, 58,* 97–105. doi: 10.1037/h0047507
4. Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 1118–1133. doi: 10.1037/a0019902
5. Butler, A. C., Karpicke, J. D., & Roediger, H. L., III. (2008). Correcting a metacognitive error: Feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 918–928. doi: 10.1037/0278-7393.34.4.918
6. Butler, A. C., & Roediger, H. L., III. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition, 36,* 604–616. doi: 10.3758/MC.36.3.604
7. Carpenter, S. K. (2012). Testing enhances the transfer of learning. *Current Directions in Psychological Science, 21,* 279–283.
8. Chan, J. C. K., McDermott, K. B., & Roediger, H. L., III. (2006). Retrieval induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General, 135,* 553–571. doi: 10.1037/0096-3445.135.4.553
9. Crowder, R. G. (1976). *Principles of learning and memory*. Hillsdale, NJ: Erlbaum.
10. Darley, C. F., & Murdock, B. B., Jr. (1971). Effects of prior free recall testing on final recall and recognition. *Journal of Experimental Psychology, 91,* 66–73. doi: 10.1037/h0031836
11. Hintzman, D. L. (2010). How does repetition affect memory? Evidence from judgments of recency. *Memory & Cognition, 38,* 102–115. doi: 10.3758/MC.38.1.102

12. Jacoby, L. L. (1974). The role of mental contiguity in memory: Registration and retrieval effects. *Journal of Verbal Learning and Verbal Behavior, 13,* 483–496. doi: 10.1016/S0022-5371(74)80001-0

13. Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30,* 513–541. doi: 10.1016/0749-596X(91)90025-F

14. Jacoby, L. L. (1999). Ironic effects of repetition: Measuring age-related differences in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 3–22. doi: 10.1037/0278-7393.25.1.3

15. Jacoby, L. L., & Wahlheim, C. N. (2013). On the importance of looking back: The role of recursive remindings in recency judgments and cued recall. *Memory & Cognition, 41,*625–637. doi: 10.3758/s13421-013-0298-5

16. Jacoby, L. L., & Wahlheim, C. N. (2014). *Memory consequences of looking back: Noticing and recollecting change*. Manuscript submitted for publication.

17. Jacoby, L. L., Wahlheim, C. N., & Coane, J. H. (2010). Test-enhanced learning of natural concepts: Effects on recognition memory, classification, and metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 1441–1451.

18. Jacoby, L. L., Wahlheim, C. N., & Yonelinas, A. P. (2013). The role of detection and recollection of change in list discrimination. *Memory & Cognition, 41,* 638–649. doi: 10.3758/s13421-013-0313-x

19. Jang, Y., & Huber, D. E. (2008). Context retrieval and context change in free recall: Recalling from long-term memory drives list isolation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 112–127. doi: 10.1037/0278-7393.34.1.112

20. Kang, S. H. K., McDaniel, M. A., & Pashler, H. (2011). Effects of testing on learning functions. *Psychonomic Bulletin & Review, 18,* 998–1005. doi: 10.3758/s13423-011-0113-x

21. Karpicke, J. D., & Roediger, H. L., III. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 704–719. doi: 10.1037/0278-7393.33.4.704

22. McDaniel, M. A., & Fisher, R. P. (1991). Tests and test feedback as learning sources. *Contemporary Educational Psychology, 16,* 192–201.

23. Nelson, S. M., Arnold, K. M., Gilmore, A. W., & McDermott, K. B. (2013). Neural signatures of test-potentiated learning in parietal cortex. *Journal of Neuroscience, 33,*11754–11762.

24. Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms [Database]. Retrieved from http://w3.usf.edu/FreeAssociation/

25. Nunes, L. D., & Weinstein, Y. (2012). Testing improves true recall and protects against the build-up of proactive interference without increasing false recall. *Memory, 20,* 138–154.

26. Otero, J., & Kintsch, W. (1992). Failures to detect contradictions in a text: What readers believe versus what they read. *Psychological Science, 3,* 229–235. doi: 10.1111/j.1467-9280.1992.tb00034.x

27. Pastötter, B., Schicker, S., Niedernhuber, J., & Bäuml, K.-H. T. (2011). Retrieval during learning facilitates subsequent memory encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 287–297. doi: 10.1037/a0021801

28. Postman, L., & Underwood, B. J. (1973). Critical issues in interference theory. *Memory & Cognition, 1,* 19–40.

29. Putnam, A. L., Wahlheim, C. N., & Jacoby, L. L. (2014). Memory for flip-flopping: Detection and recollection of political contradictions. *Memory & Cognition*. doi: 10.3758/s13421-014-0419-9

30. Radvansky, G. A., & Zacks, R. T. (1991). Mental models and the fan effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 940–953. doi: 10.1037/0278-7393.17.5.940

31. Roediger, H. L., III, & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15,* 20–27. doi: 10.1016/j.tics.2010.09.003

32. Roediger, H. L., III, & Karpicke, J. D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1,* 181–210. doi: 10.1111/j.1745-6916.2006.00012.x

33. Roediger, H. L., III, & Karpicke, J. D. (2006b). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17,* 249–255. doi: 10.1111/j.1467-9280.2006.01693.x

34. Szpunar, K. K., McDermott, K. B., & Roediger, H. L., III. (2008). Testing during study insulates against the buildup of proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 1392–1399. doi: 10.1037/a0013082

35. Tulving, E., & Watkins, M. J. (1974). On negative transfer: Effects of testing one list on the recall of another. *Journal of Verbal Learning and Verbal Behavior, 13,* 181–193. doi: 10.1016/S0022-5371(74)80043-5

36. Tzeng, O. J., & Cotton, B. (1980). A study-phase retrieval model of temporal coding. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 705–716. doi: 10.1037/0278-7393.6.6.705

37. Underwood, B. J., & Ekstrand, B. R. (1967). Studies of distributed practice: XXIV. Differentiation and proactive inhibition. *Journal of Experimental Psychology, 74*, 574–580.

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

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

40. Wahlheim, C. N., Maddox, G. B., & Jacoby, L. L. (2014). The role of reminding in the effects of spaced repetitions on cued recall: Sufficient but not necessary. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 94–105. doi: 10.1037/a0034055

41. Wegner, D. M. (1994). Ironic processes of mental control. *Psychological Review, 101*, 34–52. doi: 10.1037/0033-295X.101.1.34

42. Weinstein, Y., Gilmore, A. W., Szpunar, K. K., & McDermott, K. B. (2014). The role of test expectancy in the build-up of proactive interference in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 1039–1048. doi: 10.1037/a0036164

43. Winograd, E., & Soloway, R. M. (1985). Reminding as a basis for temporal judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 262–271. doi: 10.1037/0278-7393.11.2.262

44. Wissman, K. T., Rawson, K. A., & Pyc, M. A. (2011). The interim test effect: Testing prior material can facilitate the learning of new material. *Psychonomic Bulletin & Review, 18,*1140–1147. doi: 10.3758/s13423-011-0140-7