4-19-2016

Intermittent Testing Reduces Proactive Interference In Multiple Document Comprehension

Rebecca Marie McCabe

INTERMITTENT TESTING REDUCES PROACTIVE INTERFERENCE IN MULTIPLE DOCUMENT COMPREHENSION

by

Rebecca M. McCabe

A Thesis

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

Major: Psychology

The University of Memphis

May 2016

Abstract

The current research examined whether proactive interference (PI)—when old knowledge interferes with the learning of, and memory for, new knowledge—occurred when reading several argumentative texts on the same topic. We also examined whether retrieval practice could successfully reduce PI in this setting. In Experiment 1, participants read either eight or 24 texts on the same topic; some who read 24 texts completed retrieval practice on the first 16. All participants completed a distractor task, then a final free recall of the texts that they read. Experiment 2 explicitly measured memory for supporting evidence and sources, altered the final recall task to assess prior-text intrusions, and added a condition to rule out fatigue. Across both experiments, analyses suggest that PI occurs in a multiple document context and that this finding is not due to fatigue. Additionally, PI in this context can be reduced—if not eliminated—through retrieval practice.

Table of Contents

Intermittent Testing Reduces Proactive Interference in Multiple Document Comprehension

In daily life, readers access a wide variety of information sources including newspapers, magazines, Internet articles, blogs, and television programs. Given such diversity, successful comprehension requires that readers are able to understand and remember what each document conveys, while also maintaining and remembering the information sources themselves (i.e., which document stated which particular piece of content). Most multiple document comprehension studies have focused on the identification of reader characteristics that successfully predict comprehension such as prior knowledge (Bråten, Strømsø, & Britt, 2009; Strømsø, Bråten, & Samuelstuen, 2008) and epistemic beliefs (Bråten, Ferguson, Strømsø, & Anmarkrud, 2013). Research varying the conceptual make-up and amount of documents readers might come across, however, has been under-examined as of yet. In the current work, we contend that such a focus might elucidate mechanisms for multiple document forgetting otherwise unspecified by previous experiments using fairly small sets of documents. Connections between research in multiple document comprehension and classic memory paradigms are made for the purposes of positing a new mechanism for multiple document forgetting: *proactive interference* (PI).

PI occurs when old knowledge accumulates and increases the forgetting of new knowledge (Keppel & Underwood, 1962; Underwood, 1957). In one typical experimental method of testing the effects of PI, participants are given lists of words to memorize. After each list is learned, participants recall the items on that list. Recall for each successive list declines as more and more lists are learned. For example, Wickens (1972) instructed participants to study and recall a list of fruits. Afterwards, they studied and recalled a second, third, and fourth list of fruits. The results showed a decline in memory from list one to list four, with memory for the

fruits in the first list impairing learning and memory of fruits in subsequent lists.

There are several factors that contribute to the degree of PI that one will experience. The more potentially interfering items that need to be remembered, the more PI one will experience (Keppel & Underwood, 1962). Additionally, the more conceptual overlap between the items being learned, the more PI will occur (Postman & Keppel, 1977; Wickens, 1970, 1972; Wickens, Born, & Allen, 1963). For example, more PI occurs when items being learned are from the same category, compared to a different category (e.g., Wickens et al., 1963).

The above-mentioned research contributes to a long history of testing PI using simple materials (e.g., lists of words, nonsense syllables, word pairs). Very few studies have used more complex materials such as documents. One such example examined the existence of PI in learning about hydraulic brakes from a short expository text, situated within a larger experiment on interference in multimedia learning (Mayer, Deeleeuw, & Ayres, 2007). All participants learned about hydraulic brakes, and some learned about caliper brakes and air brakes beforehand. Those who learned about all three brake types showed poorer memory of how hydraulic brakes work, and included more intruding information about other brake types, than those who learned only about hydraulic brakes.

In the present research, we continued to examine the application of decades of PI research that used simple materials to this real-world context of reading to remember multiple documents. Specifically, we examined if PI occurs when reading a series of documents to develop a coherent understanding of several argumentative texts on a topic. Memory for a target set of eight documents (i.e., a "no-PI" control condition) was compared to a condition where PI presumably was stronger—reading to remember eight target documents after having already read 16 non-target documents. In alignment with previous research tasking individuals with

remembering lists of words, we expected multiple document comprehension to also succumb to PI. That is, individuals in the condition reading 24 documents should display poorer memory for the final set of target documents compared to the no-PI control condition, which *only* read the "final" set of target documents.

Beyond demonstrating that PI occurs when readers attempt to remember multiple documents, we additionally sought to test ways to protect against the negative effects of PI for memory in this context. Previous research points to a potential way to reduce PI through intermittent testing of the learned information (e.g., retrieval practice). The benefit of retrieval practice on memory in general is known as the *testing effect*, where testing during study leads to better recall of the practiced information on a later test (see Roediger & Butler, 2011), more so than merely re-studying the information (e.g., Roediger & Karpicke, 2006b). This benefit of testing has been shown to help memory for paired-associates (e.g., Karpicke & Roediger, 2008; Pyc & Rawson, 2009) and lists of words (e.g., Karpicke & Roediger, 2007), as well as more complex materials like texts (e.g., Karpicke & Roediger, 2010; McDaniel, Howard, & Einstein, 2009; Roediger & Karpicke, 2006a; Wissman, Rawson, & Pyc, 2011).

Of particular interest for the current work are a number of studies that suggest that one of the reasons that retrieval practice facilitates learning is that it is protective against the effects of PI (e.g., Darley & Murdock, 1971; Szpunar, McDermott, & Roediger, 2007; Tulving & Watkins, 1974). In the first study to directly test this relationship, Szpunar, McDermott, and Roediger (2008) showed that retrieval practice can indeed help to reduce PI. Participants completed five trials during which they saw a list of 18 words, did a 1-min task to prevent rehearsal, and then either completed a free recall task of the list items (i.e., retrieval practice) or performed a second rehearsal-preventing task, depending on condition. After a 30-min retention period, during which

they performed an unrelated task, all participants completed a final cumulative free recall test on the items from Lists 1-5. The results showed that completing an initial test for each list led to more correct recall of words on the final test than did merely studying the list items, which is consistent with prior research on testing in general. Additionally, interpolating retrieval practice between study of Lists 1-4 predicted better recall for List 5 items during both the initial test of List 5 and the final cumulative recall test; this effect was observed regardless of whether the words that made up the lists were conceptually related or unrelated to each other.

This finding, that retrieval practice not only benefits memory for the practiced information but also for new information learned *after* retrieval practice (thus counteracting the effects of PI), has since been replicated by others (Bäuml & Kliegl, 2013; Weinstein, McDermott, & Szpunar, 2011). In addition to the replication, Bäuml and Kliegl (2013) also showed that testing initial lists before studying the final target list leads to shorter response latencies (faster responses) during final recall of the target list (as short as participants who learned no lists prior to the target list). Similar to previous research on PI, however, the research that connects retrieval practice to PI has used simple experimental materials (i.e., lists of words, face-name pairs). In the present research, we extend this previous work to examine whether intermittent testing, functioning as opportunities to recall what a reader has comprehended thus far, reduces (if not eliminates) PI-related memory deficits. Based on prior research (Bäuml & Kliegl, 2013; Szpunar et al., 2008), we expected that those tasked to intermittently practice retrieving the texts would display better memory for information from the texts relative to those in the PI condition who did not practice retrieval and that their performance would near that of the no-PI control condition.

Experiment 1

The current study sought to better understand the most effective ways to interact with multiple documents in daily life, through expanding on previous research on PI and retrieval practice. This study addressed two research questions. Our first question was whether PI displays similar detriments in a real-world, complex document comprehension context when reading a series of argumentative texts on the same topic. We examined this question by comparing participants' memory for a target set of texts across a no-PI control condition (who *only* read that target set) and a condition that read 16 documents before reading the target set. We hypothesized that participants who read only the target set of documents would have better memory for those documents compared to those who read 16 documents before reading the target set.

Our second research question was whether retrieval practice helps to reduce PI's detrimental memory effects in the context of reading multiple argumentative texts. We examined this question by comparing participants' memory for a target set of documents across a condition that practiced retrieving the 16 documents read prior to the target set with the previously mentioned condition that did not practice retrieval of the 16 documents read prior to the target set. We hypothesized that participants who completed retrieval practice of earlier documents would be better able to recall the target set compared to those who did not have retrieval practice.

We additionally examined memory data produced specifically by the retrieval practice condition. The goal of these additional analyses was to investigate whether individual differences in retrieval practice *success* predicted memory on the final, cumulative test. We hypothesized that better performance during retrieval practice of earlier-learned texts would be related not only to better final recall of the *practiced* texts, but also to better final recall of the later-learned *unpracticed* texts. Thus, we incorporated both experimental and differential psychological

approaches (Cronbach, 1957; Underwood, 1975) to provide a more comprehensive examination concerning the role that intermittent testing might play in reducing proactive interference in multiple document comprehension.

Method

Participants. One hundred eight participants (69.4% female), with ages ranging from 18-50 (M = 22.30, SD = 6.15), participated in this study for course credit. All participants were college undergraduates recruited through the psychology subject pool of a large public university in the mid-south United States.

Materials and procedure. We randomly assigned participants to one of three conditions: PI, retrieval practice, or no-PI (control). In the *PI* condition, participants read three sets of eight documents (Sets A, B, and C), completed a distractor task, and then free recalled the 24 documents. In the *retrieval practice* condition, participants read the first set of eight documents (Set A), completed a distractor task, and then practiced retrieving those eight documents through free recall. They next read the second set of eight documents (Set B), completed a distractor task, and then practiced retrieving those eight documents through free recall. Lastly they read the third set of eight documents (Set C), completed a distractor task, and then free recalled all 24 documents. In the *no-PI (control)* condition, participants only read the final set of eight documents (Set C), completed a distractor task, and then free recalled those eight documents.

The first task completed by all participants asked them to use their own understanding to list the reasons that social media is beneficial or detrimental for society. The number of reasonable reasons that they listed served as a measure of their prior knowledge.

The 24 argumentative texts were based on information from procon.org ("Social Networking," n.d.) on the benefits and detriments of social media for society (12 of each; see the

Appendix for the eight texts that appeared in Set C). Benefits included staying connected, finding jobs, and fast spread of information; detriments included waste of time, cyber bullying, and spread of false information. Each document began with a claim statement (e.g., "Social media networking sites are beneficial for society because they spread information faster than any other media"), followed by one to three (M = 1.71) evidence statements that backed up the claim (e.g., "For example, Twitter and YouTube users reported the 2012 Aurora, Colorado theater shooting before news crews could arrive on the scene"). The body of the texts contained 58 words on average (SD = 7.38). Each document had three source features listed above the text: author occupation (e.g., journalist, social worker), publication venue (e.g., *Chicago Tribune*, www.monster.com), and publication type (e.g., journal article, news report). The 24 documents were separated into three sets of eight (Sets A, B, and C), with the same eight documents always appearing within the same set but in varying orders. Each set was presented one at a time with no ability to go back and re-read previous sets.

When finished reading, participants received a demographics questionnaire that asked for their age, gender, and GPA. Next they completed a vocabulary task that served as both an individual difference measure and a distractor task. We randomly chose 15 words (e.g., belligerent, subordinate, verbose) from a 30-item vocabulary quiz that has been normed for a college population and is correlated with reading ability (r = .52; Raney, Therriault, & Minkoff, 2000). Participants chose the best definition for each of the 15 words, each with four possible answer choices. Of the 15 unused items from the original 30-item quiz, seven were used as the distractor before retrieval practice of Set A, and seven were used as the distractor before retrieval practice of Set B in the retrieval practice condition.

Participants next completed an assessment of their memory for the documents' main

claims. Instructions were to list the benefits and detriments of social media from the documents they read. We scored participants' recall of main claims using a coding system where for each claim a score of 0, 1, or 2 could be achieved. For example, when recalling the claim statement "Social media networking sites are beneficial for society because they spread information faster than any other media," a participant would receive zero points for failing to mention the claim, one point for mentioning social media's use for spreading information, and the full two points for mentioning social media's use for spreading information *faster than other media.*

We also scored participants' total elaborations on the main claims by awarding one point for each mention of a key piece of information that was read in the evidence statements of the documents. For example, one document includes the evidence statement "Seniors report feeling happier due to online contact with family and their church community." A participant would receive one elaboration point for mentioning each of the following three key pieces of information: (1) seniors feel happier, (2) seniors have contact with family, and (3) seniors have contact with church community. Thus, a participant who mentioned all three would receive three elaboration points. Two raters independently scored a randomly selected 20% of participants' responses (Cohen's kappa = .76). Disagreements were discussed and resolved, and one rater scored the rest of the responses.

Results

Prior to addressing the research questions for this study, we established the comparability of reader characteristics across conditions. We found no significant differences between conditions for age, GPA, vocabulary, or prior knowledge. Below, we first present our analyses of the final recall across the three conditions. We then present our analyses of recall performance within the retrieval practice condition (see Table 1 for all means, standard deviations, and test

statistics).[1]

Final recall. We first examined the three conditions for differences in memory for the claims of the last eight documents (the target set). We hypothesized that participants who read only the target set of documents would have better memory for those documents compared to those who read 16 documents before reading the target set, and also participants who practiced retrieval of the first 16 documents would have better memory of the target set than those who did not practice retrieval. A one-way between-groups analysis of variance found a significant difference, $F(2, 102) = 12.26$, $p < .001$, $\eta^2 = .19$. In regards to the simple effects, the mean memory score (with higher numbers indicating better memory) for the no-PI condition ($M = 9.72$, $SD = 2.06$) and the retrieval practice condition ($M = 8.47$, $SD = 2.73$) were both significantly greater than the PI condition ($M = 6.76$, $SD = 2.63$), consistent with our hypotheses. Also, the no-PI condition was significantly higher than the retrieval practice condition. Thus, the PI condition showed the most PI, and while those in the retrieval practice condition appear to have benefitted from the intermittent testing, they did not remember as much as the no-PI condition who had no opportunity for PI to occur.

We next examined the three conditions for differences in the number of elaborations from the evidence statements of the last eight documents (the target set) that they included during final free recall. Because our data did not pass Levene's test for homogeneity of variance, we used the Welch's adjusted F ratio, and found a significant difference between conditions, $F(2, 63.83) = 3.84$, $p = .027$, $\eta^2 = .08$. Post-hoc comparisons using the Games-Howell test revealed that the mean elaboration score for the no-PI condition ($M = 6.22$, $SD = 5.94$) was significantly greater than the PI condition ($M = 3.16$, $SD = 3.41$), consistent with our hypothesis. The retrieval

[1] We additionally ran all of the presented analyses using ANCOVA or partial correlations to rule out any effect of prior knowledge or vocabulary. The results of these analyses do not differ from what we present here. Thus, in the interest of simplicity, they are not reported.

practice condition (M = 3.10, SD = 5.00) did not differ significantly from the no-PI or PI conditions. We must be careful in interpreting these results, however, because participants did not receive explicit instructions to include elaborations while listing the benefits and detriments of social media. Therefore, a participant not including elaborations is not necessarily indicative of their lack of memory for the texts' evidence statements. Thus, what we can conclude is that participants in the no-PI condition decided to include more elaborations than participants in the PI condition irrespective of the fact that they were not explicitly asked to do so.

Recall during retrieval practice. In order to more specifically examine the relationships between memory for the different document sets for readers in the retrieval practice conditions, we explored potential relationships between individual differences in retrieval practice success and final cumulative recall success. The hypothesized positive correlation between retrieval practice of the first 16 documents and final recall of the first 16 documents was significant. This was true for both memory for the claims (r = .773, p < .001) and elaborations (r_s = .711, p < .001). Thus, successful retrieval of earlier-read documents during practice was associated with successful retrieval of these documents during the final cumulative recall test.

Additionally, the hypothesized positive correlation between retrieval practice of the first 16 documents and final recall of the target final eight documents was also significant. Again, this was true for both memory for the claims (r = .346, p = .039) and elaborations (r_s = .686, p < .001). Thus, successful retrieval practice of earlier-read documents was associated with improved memory for the final unpracticed document set (i.e., the documents that should have experienced the most proactive interference), again showing how retrieval practice can insulate against PI.

Discussion

The results of Experiment 1 suggest that PI occurs in a multiple document context and that its detrimental effects can be reduced (but not eliminated) in this context through intermittent testing. Participants who experienced no PI recalled the most target text claims during final recall, whereas participants who experienced the most PI recalled the least. Participants who performed retrieval practice recalled more claims than the participants who experienced the most PI, but not as much as participants who experienced no PI, which suggests that retrieval practice reduces but does not eliminate PI in the context of learning from multiple documents.

Additionally, better retrieval practice performance was related to better memory for the unpracticed target texts during final recall. This is consistent with the finding that retrieval practice of earlier-learned information can lead to better memory for later-learned information (Bäuml & Kliegl, 2013; Szpunar et al., 2008; Weinstein et al., 2011), and it again suggests that retrieval practice can guard readers against PI. Also, better retrieval practice performance was related to better memory for the practiced documents during final recall, consistent with previous research (see Roediger & Karpicke, 2006b).

However, there are limitations to Experiment 1 that leave unanswered questions. In Experiment 2, we sought to address three limitations. First, one could argue that Experiment 1's PI condition recalled fewer arguments from the final 8 texts simply due to fatigue and not necessarily due to increased interference from the content of the previously-read documents. While this explanation seemed unlikely, considering that the retrieval practice condition also read 24 texts and still performed better than the PI condition during final recall, we sought to rule out this interpretation by adding a fourth condition: a *release from PI* condition.

In studying memory for lists of words, Wickens (1972) demonstrated that release from PI occurs when the items learned in the final list have little conceptual overlap with items in the previous lists. Moreover, this release from PI appears to be a function of the degree of conceptual overlap. In his study, all participants first studied and recalled three lists of fruits, with recall performance declining with each successive list. For the fourth list, depending on their condition, participants studied and recalled either a list of fruits, vegetables, or professions. Recall performance for the fourth list of fruits continued to decline, whereas recall performance on the list of vegetables was only slightly better (due to there being some conceptual overlap between fruits and vegetables), and recall performance on the list of professions was nearly as good as the first list of fruits (due to the lack of conceptual overlap between fruits and professions).

Similarly, Experiment 2 examined if release from PI occurs in the context of reading multiple complex texts as a way to rule out fatigue as an explanation for the results of Experiment 1. Participants in the new release from PI condition read 16 texts on an unrelated topic (reasons why *standardized testing* is beneficial or detrimental for society) before reading the target eight texts on reasons why *social media* is beneficial or detrimental for society. We hypothesized that participants who first read 16 texts that *were* conceptually related to the target eight texts (PI condition) would remember less from the target texts than participants who first read 16 texts that *were not* conceptually related to the target eight texts (release from PI condition). Additionally, we hypothesized comparable memory for the release and no-PI conditions, which would suggest that fatigue was not the mechanism for forgetting in the PI condition.

The second limitation of Experiment 1 was that the final recall test did not permit us to assess what has often been used in previous PI research: *prior-list intrusions* (Zaromb et al.,

2006). Prior-list intrusions occur when participants learn items from multiple lists, and when they are asked to recall items from *only* the final target list, they mistakenly include items from previous lists as well. Thus, items learned from prior lists intrude into their recall of the target list. Prior-list intrusions are used as an additional indication of PI (e.g., Szpunar et al., 2008), with previous research showing that experiencing more PI leads to more prior-list intrusions (e.g., Rosen & Engle, 1998). In the case of the current research, these might be referred to as prior-*set* intrusions because participants are learning from sets of texts instead of lists of words.

The way Experiment 1 was designed, we were unable to measure prior-set intrusions because we asked participants to recall items from *all* sets during final recall. In Experiment 2, however, we asked participants to recall items from only the final target set of eight texts. This allowed us to both measure prior-set intrusions to use as further evidence of PI and to align our methods with that of previous research. We hypothesized that participants who read more texts that were conceptually related, and who did not receive retrieval practice, would include more prior-set intrusions during final recall.

Finally, the results from Experiment 1 were limited by our use of shallow assessment of participants' memory for the texts. That is, we only asked them to list the benefits and detriments of social media that they remembered reading. The task instructions were vague enough that participants may have only thought it necessary to include texts' claims. Thus, if a participant did not include information from texts' evidence statements during final recall, we could not conclude that they did not remember the evidence statements per se. They may have thought it was unnecessary to include them, or they lacked the motivation to do so. To address this issue, and to dig deeper into assessing participants' learning of the texts, Experiment 2 included two new dependent measures to explicitly prompt participants' memory for the texts' evidence

statements and sources. We hypothesized that participants who read more texts that were conceptually related, and who did not receive retrieval practice, would remember less evidence statements from the texts. Additionally, because they would not receive an opportunity to practice retrieval on texts' sources, we hypothesized that memory for texts' authors would not be different between the retrieval practice condition and the PI condition.

Thus, in summary, Experiment 2 addressed these three limitations through the following three goals: (1) to replicate the differences in claim memory across the PI, retrieval practice, and no-PI conditions while using prior-set intrusions as an additional measure of PI; (2) to rule out fatigue as an explanation of results through the addition of a release from PI condition; (3) to test for additional detrimental effects of PI and additional benefits of retrieval practice guarding against PI by also measuring participants' memory for additional facets of argumentative texts, namely the evidence provided and the sources who wrote the arguments.

Experiment 2

Method

Participants. A total of 85 participants (67.1% female), with ages ranging from 18 to 47 (M = 20.93, SD = 4.30), participated in this study for course credit. All participants were college undergraduates recruited through the psychology subject pool of a large public university in the mid-south United States.

Materials and procedure. We randomly assigned participants to one of four conditions: PI, retrieval practice, no-PI (control), or release from PI. All aspects of the materials and procedure in Experiment 1 remained the same in Experiment 2 except for the following: the addition of the release from PI condition, the addition of measures of memory for text evidence and sources, and the claim recall task was changed to be recall of only the target set of texts (Set

C). Final recall responses were scored the same way as in Experiment 1 (Cohen's kappa = .84).

Contrasted with the other three conditions, the release from PI condition first read two sets of eight texts on an unrelated topic (i.e., the benefits and detriments of standardized testing for society; Sets A*, B*), followed by the target set of eight texts on the benefits and detriments of social media for society (Set C). We developed the 16 standardized testing texts the same way, and with the same format, as the social media texts. That is, all texts have a main claim, 1-3 evidence statements, and three source features (author occupation, publication venue, and publication type).

The evidence memory assessment asked participants to identify whether or not they remembered reading a given evidence sentence in the last set of texts that they read (Set C). The measure consisted of 16 total test sentences, including a randomized order of eight sentences from texts in Set C, four sentences from texts in Sets A, and four sentences from texts in Set B. After each sentence, participants were asked to circle *yes* if they saw the sentence in the final set of eight texts that they read or *no* if they did not. As an exploratory measure, they were then asked to choose a number 1-5 to indicate how confident they were in their response. This same format was used for the measure of memory for sources.

Results

Prior to addressing the research questions for Experiment 2, we established the comparability of reader characteristics across the four conditions. We found no significant differences between conditions for age, GPA, vocabulary, or prior knowledge. Below, we first present our analyses of the final recall across the four conditions. We next present our analyses on the evidence and source memory tasks and the confidence ratings given on these tasks. Finally, we present our analyses of recall performance within the retrieval practice condition (see

Table 2 for all means, standard deviations, and test statistics).[2]

Final claim recall. Patterns associated with the negative effect of PI on memory include a decrease in accurate responses and an increase in number of intrusions from previously learned information. As such, when scoring final claim recall, we created an adjusted composite score that accounts for these two facets. The combination was such that we subtracted the number of prior-set intrusions from the number of accurate claims recalled (hereafter referred to as claim recall accuracy; Postman & Keppel, 1977).

A one-way Welch's ANOVA examined the four conditions for differences in adjusted claim recall accuracy from the last eight documents (see Figure 1). The results were significant, Welch's $F(3, 44.02) = 45.54$, $p < .001$, $\eta^2 = .61$. We hypothesized that the release condition would have better accuracy than the PI condition and similar accuracy as the control condition. This hypothesis was partially supported. A Games-Howell post-hoc test revealed no significant difference between the retrieval practice ($M = 9.12$, $SD = 3.03$) and control ($M = 10.14$, $SD = 2.20$) conditions, suggesting similarly accurate recall across the two conditions. The release condition ($M = 7.04$, $SD = 3.76$) was not significantly different from the retrieval practice condition, but the release condition was significantly less accurate than the control condition. The PI condition ($M = 0.27$, $SD = 3.31$) was significantly less accurate than all other conditions, mistakenly recalling nearly as many intrusions as they did correct claims. Thus, the control and retrieval practice conditions performed the best, the PI condition performed the worst, and the release condition performed between the control and PI conditions.

Evidence memory. In our analysis of evidence memory, we first used a one-way ANOVA to examine the four conditions for differences in their performance on the evidence

[2] As in Experiment 1, we additionally ran all of the presented analyses using ANCOVA or partial correlations to rule out any effect of prior knowledge or vocabulary. The results of these analyses do not differ from what we present here. Thus, in the interest of simplicity, they are not reported.

memory task, specifically on items that pertained to evidence read in the final set of eight texts. The results were significant, $F(3, 80) = 5.67$, $p = .001$, $\eta^2 = .18$. Our hypothesis—that reading 24 texts on the same topic would lead to poor memory for supporting evidence—was supported. A post-hoc test Tukey HSD test revealed that the PI condition ($M = 5.77$, $SD = 1.77$) was less able to recognize which evidence statements they had read in the final set of eight texts than all other conditions (control: $M = 7.43$, $SD = 0.68$; release: $M = 6.82$, $SD = 1.44$; retrieval practice: $M = 6.79$, $SD = 1.18$). No other effects reached statistical significance.

We next used an independent-samples t-test to examine the PI and retrieval practice conditions for differences in the number of prior-set intrusions they had during the evidence memory task (i.e., how many items from the first 16 texts they mistakenly identified as coming from the last eight texts). The control and release conditions were excluded from this analysis because it was not possible for them to experience prior-set intrusions of evidence statements due to the fact that they only read the final set of eight target texts on the benefits and detriments of social media. The results of the t-test were significant, $t(28.48) = 2.16$, $p < .05$, $\eta^2 = .11$. The PI condition ($M = 5.27$, $SD = 2.23$) had significantly more prior-set intrusions during the evidence memory task than the retrieval practice condition ($M = 3.28$, $SD = 3.36$).

Source memory. In our analysis of source memory, we first used a one-way ANOVA to examine the four conditions for differences in their performance on the source memory task, specifically on items that pertained to sources from the final set of eight texts. The results were significant, $F(3, 80) = 10.54$, $p < .001$, $\eta^2 = .28$. Our hypothesis—that the retrieval practice and PI conditions would experience similarly low levels of source memory—was not supported. A post-hoc Tukey HSD test revealed that the retrieval practice condition ($M = 4.63$, $SD = 1.71$) performed significantly worse than all other conditions (PI: $M = 5.95$, $SD = 1.46$; release: $M =$

6.05, *SD* = 1.40; control: *M* = 7.19, *SD* = 1.17), meaning that they were the least accurate in recognizing which sources they had read in the final set of eight texts. Additionally, the PI condition performed significantly worse than the control condition. No other effects reached statistical significance.

We next used a one-way ANOVA to examine the PI, retrieval practice, and release conditions for differences in the number of prior-set intrusions they had during the source memory task (i.e., how many sources from the first 16 texts they mistakenly identified as coming from the last eight texts). The control condition was excluded from this analysis because it was not possible for them to experience prior-set intrusions of sources due to the fact that they only read the final set of eight target texts. The results of the ANOVA were significant, $F(1, 61) = 8.00$, $p = .001$, $\eta^2 = .21$. A post-hoc Tukey HSD test revealed that retrieval practice condition (*M* = 3.00, *SD* = 1.49) had significantly less prior-set intrusions during the source memory task than both the PI (*M* = 4.95, *SD* = 1.56) and release (*M* = 4.74, *SD* = 1.96) conditions. No other effects reached statistical significance.

Confidence ratings. We first explored confidence ratings for the evidence memory task. Across the four conditions, a one-way ANOVA compared participants' average confidence rating of items on the evidence memory task that pertained to the final set of eight target texts. The ANOVA was not significant, suggesting no difference between conditions in how confident they were in their ability to recognize supporting evidence that came from the final set of eight texts.

We next explored confidence ratings for the source memory task. Across the four conditions, a one-way ANOVA compared participants' average confidence rating of items on the source memory task that pertained to the final set of eight target texts. The ANOVA was

significant, $F(3, 78) = 11.63$, $p < .001$, $\eta^2 = .31$. A post-hoc Tukey HSD test revealed that the average confidence ratings of the control condition ($M = 4.69$, $SD = .37$) were significantly higher than all other conditions (retrieval practice: $M = 3.76$, $SD = .55$; PI: $M = 4.01$, $SD = .52$; release: $M = 4.27$, $SD = .59$). Additionally, the release condition was significantly more confident than the retrieval practice condition. No other effects reached statistical significance. These results indicated that participants who only read the final eight texts had the most confidence in their ability to recognize the sources of those texts. Participants who read 16 texts before the final target set of eight texts were more confident in their ability to recognize the sources if the first 16 texts were on an unrelated topic than if they had the opportunity to practice retrieval of them.

Recall during retrieval practice. As in Experiment 1, we specifically examined the relationships between memory for the different document sets for readers in the retrieval practice condition. Using Pearson product-moment correlation coefficients, we explored potential relationships between individual differences in retrieval practice performance and final cumulative recall performance. Consistent with the results of Experiment 1, the hypothesized positive correlation between retrieval practice of the first 16 documents and final recall of the target final eight documents was significant. This was true for both memory for the claims ($r = .503$, $p = .02$) and elaborations ($r = .731$, $p < .001$). Thus, successful retrieval practice of earlier-read documents was associated with improved memory for the final unpracticed document set (i.e., the documents that should have experienced the most proactive interference). As in Experiment 1, this suggests that retrieval practice can insulate against PI.

General Discussion

Every day, readers come across a wide array of information sources, including Internet

articles, newspapers, magazines, and blogs. Because of this diversity, successful comprehension requires that readers are able to understand and remember the content itself, while also remembering which document stated which particular piece of content. In the current research, we contend that a focus on the conceptual make-up and amount of documents that readers come across can aid our understandings of the cognitive mechanisms of multiple document forgetting.

The purpose of the current research was to examine whether proactive interference (PI)—when old knowledge interferes with the learning of, and memory for, new knowledge—occurred in the context of learning from multiple documents, specifically when reading a series of argumentative texts on the same topic. A secondary purpose was to examine if retrieval practice could successfully reduce PI in this context. In Experiment 1, participants read either eight or 24 texts on the same topic. Some participants who read 24 texts completed retrieval practice on the first 16. Participants completed a distractor task, then a final free recall of the texts that they had read. Experiment 2 added explicit measures of memory for supporting evidence and sources, altered the final recall task to assess intrusions from prior texts, and included an additional condition to rule out fatigue as an alternative explanation. In this new condition, participants read 24 texts, but the first 16 texts were on a different topic from the final eight. Participants' memory for the claims, supporting evidence, and sources of the final eight texts were compared across conditions. Overall, the findings from Experiment 2 replicated and extended those from Experiment 1.

PI in Multiple Texts

As hypothesized, the results of Experiments 1 and 2 indicated that reading a series of 24 related argumentative texts led to worse memory for the final eight texts in the series, compared to only reading eight texts in total. Consistent with our hypotheses, this pattern of worse memory

occurred for the texts' claims, supporting evidence, and sources. Together, these findings suggest that PI occurs when learning from multiple texts, extending previous research on the occurrence of PI when learning from simple experimental materials like lists of words (e.g., Wickens, 1972), nonsense syllables (e.g., Keppel & Underwood, 1962), and word pairs (e.g., Postman & Keppel, 1977). Additionally, participants may have been—on some level—aware of this interference, judging by their low confidence in their memory for supporting evidence and sources (ratings of confidence in the correctness of answer choices are often used as a measure of metamemory; Bornstein & Zickafoose, 1999; Nelson & Narens, 1990).

Taking into account alternative explanations of our results, Experiment 2 served to strengthen PI as the mechanism behind our results in two ways. First, it used a more strict measure of PI: prior-set intrusions. Intrusions occur when participants learn items from multiple lists or sets, and items learned from prior sets intrude into their recall of the final target set. Consistent with our hypothesis, reading 24 texts on the same topic led to worse claim accuracy (worse memory and more prior-set intrusions) during recall of the final eight texts. This finding is consistent with previous research showing that experiencing PI leads to both worse memory and more intrusions (e.g., Rosen & Engle, 1998; Szpunar et al., 2008; Zaromb et al., 2006).

Release From PI

The second way that Experiment 2 strengthened PI as the mechanism behind our results was that it ruled out fatigue as an alternative explanation using the release from PI paradigm (Wickens, 1972), where reducing the amount of conceptual overlap between sets of information reduces PI; the more disparate the concepts, the less PI is experienced. Consistent with our hypothesis, reading 16 texts on an unrelated topic before the final eight texts led to more accurate claim memory than reading all 24 texts on the same topic, which replicates the release from PI

paradigm using complex texts instead of more simple experimental materials (e.g., Wickens, 1972; Wickens et al., 1963). This finding indicates that it was not merely the act of reading 24 texts that led to poor memory accuracy; it was reading 24 texts *on the same topic*.

While those who read the 16 texts on an unrelated topic before the final eight texts did show more accurate claim memory relative to those who read all 24 texts on the same topic, they showed less accurate claim memory relative to the control condition (those who only read the final eight texts), which was contrary to our hypothesis. Taken together, these findings suggest that a minor amount of fatigue may have occurred from reading 16 additional texts, but it was not enough to account for PI-related forgetting. If fatigue were the sole explanation, one could expect to see the same performance from participants who read 24 texts, regardless of text topic.

In addition to better claim memory, better memory for supporting evidence from the final eight texts was observed when the first 16 texts were on a topic unrelated to that of the first eight texts, compared to when they were on the same topic, which was consistent with our hypothesis. This again suggests that the results of Experiment 1 were not due to fatigue. Memory for sources, however, was not affected by the relatedness of the first 16 texts. Thus, it appears that evidence memory was released from PI, while source memory was not.

Lack of release for sources may have occurred because participants did not connect the source to its content. Ideally, readers would be able to remember who said what, linking the content to its linked source. If this were the case, releasing the content from PI would also release its connected source. Previous research, however, has shown that readers do not link sources to content spontaneously, nor do they do it well (Britt & Aglinskas, 2002; Sparks & Rapp, 2011; Wiley et al., 2009). In the current research, sources not linked to their content are not unlike Wickens' (1972) list of professions—or any other list of conceptually related words—

with no conceptual shift to differentiate them from each other (e.g., senior writer, data analyst, social worker). Restated, regardless of the topic of the text itself, each text's source (i.e., the author's occupation) is essentially one item in a long list of 24 to-be-remembered professions. Unlinked to its text, a source would have been unlinked to the conceptual shift during the topic change, thus leaving the sources unreleased from PI.

Retrieval Practice

In addition to examining whether PI occurs in the context of learning from multiple texts, a second purpose of the current research was to examine if retrieval practice could help to reduce PI in this context, as it has been shown to do in the simpler context of learning multiple lists of words (Szpunar et al., 2008). Consistent with our hypotheses, this idea was supported by the results of Experiments 1 and 2: Practicing retrieval of the first 16 texts led to better memory for claims from the final target set of eight texts. Taken together, Experiments 1 and 2 suggest that retrieval practice can help to reduce—if not eliminate—PI when learning from multiple texts, consistent with previous research using lists of words (Szpunar et al., 2008).

Compared to simply reading the 24 texts, an opportunity to practice retrieval of the first 16 texts also led to better memory for supporting evidence from the final eight texts (consistent with our hypothesis), whereas it actually hurt memory for the texts' sources. As hypothesized, however, retrieval practice did lead to fewer prior-set intrusions during both the evidence and source memory tasks compared to participants who simply read the 24 texts (though it did not affect response confidence). This is consistent with the idea that retrieval practice helps differentiate what was learned recently (in the final set of eight texts) from what was learned earlier (Szpunar et al., 2008).

Our finding that retrieval practice hurt memory for sources was likely due to the fact that

the retrieval practice was a free recall task where participants listed what they could remember reading. All participants practiced retrieval of the texts' claims, most practiced the supporting evidence, and almost none practiced the sources. Perhaps this was because the instructions did not explicitly ask participants to recall sources, or perhaps it was because participants did not attend to the sources in the first place—an occurrence all too common (Britt & Aglinskas, 2002; Sparks & Rapp, 2011). In either case, because retrieval practice of the sources was rare, it is possible that retrieval-induced forgetting (Murayama, Miyatsu, Buchli, & Storm, 2014) occurred. Retrieval-induced forgetting refers to how retrieval of a subset of studied items results in worse memory for the unpracticed items. In the case of the current research, it is possible that a similar type of trade-off occurred, whereby retrieval of the texts' claims and evidence led to poorer memory for the texts' sources. Future research could benefit from exploring the existence of retrieval-induced forgetting in this context, perhaps by giving participants explicit instructions or tasks to practice the retrieval of supporting evidence and sources.

Across both experiments, when comparing participants within the retrieval practice condition, better memory during retrieval practice was related to better memory for those same texts during final recall, which is consistent with findings on the benefits of retrieval practice (Roediger & Butler, 2011; Roediger & Karpicke, 2006b). Additionally, better memory during retrieval practice of the first 16 texts was related to better memory during final recall of the last eight texts. This finding is consistent with previous research on the forward effect of retrieval practice (Pastötter & Bäuml, 2014; Szpunar et al., 2008), which shows that retrieval practice helps memory not only for what is practiced, but also for new information that is learned after the retrieval practice event.

Several cognitive mechanisms have been suggested for why retrieval practice helps to

reduce PI. One theory is that it allows learners to better encode items during initial study, with each test allowing learners to essentially reset their encoding process for the next list. In a study by Pastötter, Schicker, Niedernhuber, and Bäuml (2011), EEG recordings during information encoding suggested an increasing lack of attention when retrieval practice did not occur between lists of information, compared to no change in attention when retrieval practice did occur between lists. Another theory is that having to complete retrieval practice after learning each set of information helps learners to maintain test expectancy, thus motivating them to maintain a certain level of attention in anticipation of upcoming tests (Weinstein, Gilmore, Szpunar, & McDermott, 2014).

While the results of the current study cannot specifically provide evidence for either of these first two theories, they are in alignment with a third theory: practicing retrieval between sets (or lists) helps the learner to discriminate between sets and thus to differentiate which items came from which set (Szpunar et al., 2008). This ultimately allows the learner to limit the size of their search set when trying to recall target information, as evidenced by how practicing retrieval on earlier learned information leads to faster response time when recalling later learned information (Bäuml & Kliegl, 2013). The results of the current study also appear to be in line with this theory. Although we cannot speak to response times, practicing retrieval between sets did lead to fewer prior-set intrusions during final memory measures, suggesting a better ability to differentiate between what was learned in the final target set of texts compared to what was learned in the first two sets.

Limitations and Future Directions

To further examine the theoretical mechanisms behind the benefit of retrieval practice for reducing PI in the context of learning from texts, future research should consider recording

response time during memory measures, with the expectation that those who practiced retrieval before learning the final target set of texts would recall information from the final target set faster than those who did not. Along the same line of inquiry, future research could also provide participants with a random order of all 24 texts at the end of the study and ask them to explicitly label which of the three sets each text came from, with the expectation that those who practiced retrieval between sets of texts would be more accurate at this tasks than those who did not.

The current research is limited in that we only examined the occurrence of PI in one type of text (argumentative) and in a specific number of texts (24). Readers could encounter many different kinds of texts (e.g., expository, narrative) and in varying numbers. Future research should consider manipulating the type of texts that are read and how many are read. The current research is also limited in that we only examined the occurrence of PI during retrieval when it has been suggested that PI also occurs during encoding (Kliegl, Pastötter, & Bäuml, 2015). Future research would benefit from including online trace methodologies (e.g., think alouds, eye tracking) to record the processing that occurs while learners interact with and encode the texts.

Future research should also utilize different encoding manipulations that have been used in the context of reducing PI from simple experimental materials to see if they can also—like retrieval practice—reduce PI in the context of learning from multiple documents. For example, one previous study used an internal context change (an imagination task) to reduce PI when learning lists of words (Pastötter & Bäuml, 2007). Similarly, future research could have participants complete an imagination task between sets of texts, with the expectation that this would lead to better recall of the final set of target texts. Another previous study used directed forgetting to reduce PI when learning lists of words (Bjork, 1970). Similarly, future research could cue participants to forget nontarget texts, with the expectation that this would lead to better

recall of the target texts.

Lastly, future research might also include individual differences measures to explore how some readers may be more susceptible to PI when reading multiple texts than others. One example would be working memory capacity, which is related to the amount of PI that one experiences when learning lists of words (Kane & Engle, 2000). Because working memory capacity is also related to reading comprehension (McVay & Kane, 2012), working memory might differentially relate to PI when learning from multiple texts compared to when learning lists of words.

Conclusion

The current series of experiments contributes uniquely to research on readers' memory for arguments on a topic presented across multiple documents. The results suggest that PI is a mechanism of multiple document forgetting (i.e., texts read early in a series are better remembered than texts read later in a series when all texts are on a related topic), specifically occurring when learning from multiple argumentative texts. The detrimental effect of PI in this context can be reduced—if not eliminated—through retrieval practice, though the results stress the importance of practicing retrieval of all to-be-remembered information and not just a subset of information. Importantly, the results indicate that this worse memory for later-read texts is not simply due to fatigue. These findings have implications for how readers interact with a series of texts as they try to learn from them. It will be important for future research to continue to explore the occurrence of PI in the everyday task of learning from texts.

References

Bäuml, K.-H. T., & Kliegl, O. (2013). The critical role of retrieval processes in release from proactive interference. *Journal of Memory and Language*, *68*, 39-53. doi:10.1016/j.jml.2012.07.006

Bjork, R. A. (1970). Positive forgetting: The noninterference of items intentionally forgotten. *Journal of Verbal Learning and Verbal Behavior*, *9*, 255-268. doi:10.1016/S0022-5371(70)80059-7

Bornstein, B. H., & Zickafoose, D. J. (1999). "I know I know it, I know I saw it": The stability of the confidence-accuracy relationship across domains. *Journal of Experimental Psychology: Applied*, *5*, 76-88. doi:10.1037/1076-898x.5.1.76

Bråten, I., Ferguson, L. E., Strømsø, H. I., & Anmarkrud, Ø. (2013). Justification beliefs and multiple-documents comprehension. *European Journal of Psychology of Education*, *28*, 879-902. doi:10.1007/s10212-012-0145-2

Bråten, I., Strømsø, H. I., & Britt, M. A. (2009). Trust matters: Examining the role of source evaluation in students' construction of meaning within and across multiple texts. *Reading Research Quarterly*, *44*, 6-28. doi:10.1598/RRQ.44.1.1

Britt, M. A., & Aglinskas, C. (2002). Improving students' ability to identify and use source information. *Cognition and Instruction*, *20*, 485-522. doi:10.1207/S1532690XCI2004_2

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist, 12,* 671-684. doi:10.1037/h0043943

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

Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and

divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 336-358. doi:10.10371/0278-7393.26.2.336

Karpicke, J. D., & Roediger, H. L., III. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language, 57,* 151-162. doi:10.1016/j.jml.2006.09.004

Karpicke, J. D., & Roediger, H. L., III. (2008). The critical importance of retrieval for learning. *Science, 319,* 966-968. doi:10.1126/science.1152408

Karpicke, J. D., & Roediger, H. L., III. (2010). Is expanding retrieval a superior method for learning text materials? *Memory & Cognition, 38,* 116-124. doi:10.3758/MC.38.1.116

Keppel, G., & Underwood, B. J. (1962). Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior*, *1*, 153-161. doi:10.1016/S0022-5371(62)80023-1

Kliegl, O., Pastötter, B., & Bäuml, K.-H. T. (2015). The contribution of encoding and retrieval processes to proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*, 1778-1789. doi:10.1037/xlm0000096

Mayer, R. E., Deleeuw, K. E., & Ayres, P. (2007). Creating retroactive and proactive interference in multimedia learning. *Applied Cognitive Psychology*, *21*, 795-809. doi:10.1002/acp.1350

McDaniel, M. A., Howard, D., & Einstein, G. O. (2009). The read–recite–review study strategy: Effective and portable. *Psychological Science, 20,* 516-522. doi:10.1111/j.1467-9280.2009.02325.x

McVay, J. C., & Kane, M. J. (2012). Why does working memory capacity predict variation in

reading comprehension? On the influence of mind wandering and executive attention. *Journal of Experimental Psychology: General*, *141*, 302-320. doi:10.1037/a0025250

Murayama, K., Miyatsu, T., Buchli, D., & Storm, B. C. (2014). Forgetting as a consequence of retrieval: A meta-analytic review of retrieval-induced forgetting. *Psychological Bulletin*, *140*, 1383-1409. doi:10.1037/a0037505

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *The Psychology of Learning and Motivation, 26,* 125-141. doi:10.1016/s0079-7421(08)60053-5

Pastötter, B., & Bäuml, K.-H. (2007). The crucial role of postcue encoding in directed forgetting and context-dependent forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 977-982. doi:10.1037/0278-7393.33.5.977

Pastötter, B., & Bäuml, K.-H. T. (2014). Retrieval practice enhances new learning: The forward effect of testing. *Frontiers in Psychology*, *5*, 1-5. doi:10.3389/fpsyg.2014.00286

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

Postman, L., & Keppel, G. (1977). Conditions of cumulative proactive inhibition. *Journal of Experimental Psychology: General*, *106*, 376-403. doi:10.1037/0096-3445.106.4.376

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language, 60,* 437-447. doi:10.1016/j.jml.2009.01.004

Raney, G. E., Therriault, D., & Minkoff, S. (2000). Repetition effects from paraphrased text: Evidence for an integrated representation model of text representation. *Discourse*

Processes, 29, 61-81. doi:10.1207/S15326950dp2901_4

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

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

Roediger, H. L., III, & Karpicke, J. D. (2006b). 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

Rosen, V. M., & Engle, R. W. (1998). Working memory capacity and suppression. *Journal of Memory and Language*, *39*, 418-439. doi:10.1006/jmla.1998.2590

Social Networking (n.d.). Retrieved from http://socialnetworking.procon.org

Sparks, J. R., & Rapp, D. N. (2011). Readers' reliance on source credibility in the service of inference generation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 230-247. doi:10.1037/a0021331

Szpunar, K. K., McDermott, K. B., & Roediger, H. L., III. (2007). Expectation of a final cumulative test enhances long-term retention. *Memory & Cognition, 35,* 1007-1013. doi:10.3758/BF03193473

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

Strømsø, H. I., Bråten, I., & Samuelstuen, M. S. (2008). Dimensions of topic-specific epistemological beliefs as predictors of multiple text understanding. *Learning and*

Instruction, *18*, 513-527. doi:10.1016/j.learninstruc.2007.11.001

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

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

Underwood, B. J. (1975). Individual differences as a crucible in theory construction. *American Psychologist*, *30*, 128-134. doi:10.1037/h0076759

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

Weinstein, Y., McDermott, K. B., & Szpunar, K. K. (2011). Testing protects against proactive interference in face-name learning. *Psychonomic Bulletin & Review*, *18*, 518-523. doi:10.3758/s13423-011-0085-x

Wickens, D. D. (1970). Encoding categories of words: An empirical approach to meaning. *Psychological Review*, *77*, 1-15. doi:10.1037/h0028569

Wickens, D. D. (1972). Characteristics of word encoding. In A. W. Melton & E. Martin (Eds.), *Coding Processes in Human Memory* (pp. 191-215). New York, NY: Wiley.

Wickens, D. D., Born, D. G., & Allen, C. K. (1963). Proactive inhibition and item similarity in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *2*, 440-445. doi: 10.1016/S0022-5371(63)80045-6

Wiley, J., Goldman, S. R., Graesser, A. C., Sanchez, C. A., Ash, I. K., & Hemmerich, J. (2009).

Source evaluation, comprehension, and learning in internet science inquiry tasks. *American Educational Research Journal*, *46*, 1060-1106.

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

Zaromb, F. M., Howard, M. W., Dolan, E. D., Sirotin, Y. B., Tully, M., Wingfield, A., & Kahana, M. J. (2006). Temporal associations and prior-list intrusions in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 792-804. doi:10.1037/0278-7393.32.4.792

Table 1

Experiment 1: Statistical values for claim recall measures for participants in the No-PI Control (n = 36), Retrieval Practice (n = 36), and Proactive Interference (n = 36) conditions.

Measures	Condition			Test Statistic	Effect Size
	No-PI Control	Retrieval Practice	Proactive Interference		
	M (*SD*)	*M* (*SD*)	*M* (*SD*)		
Claim Recall					
Set A	N/A	5.39 (3.02)	5.28 (2.91)	$t < 1$, ns	
Set B	N/A	6.75 (3.07)	5.27 (2.71)	$t(67) = -2.11$, $p = .04$	$\eta^2 = .06$
Set C	9.72 (2.06)	8.47 (2.73)	6.76 (2.63)	$F(2, 102) = 12.26$, $p < .001$	$\eta^2 = .19$

Note. Higher scores indicate a greater magnitude of each variable.

Table 2

Experiment 2: Statistical values for claim recall measures, evidence memory measures, and source memory measures for participants in the No-PI Control (n = 21), Retrieval Practice (n = 19), Proactive Interference (n = 22), and Release (n = 23) conditions.

Measures	Condition				Test Statistic	Effect Size
	No-PI Control *M* (*SD*)	Retrieval Practice *M* (*SD*)	Proactive Interference *M* (*SD*)	Release from PI *M* (*SD*)		
Claim Recall						
Set C	10.14 (2.20)	9.12 (3.03)	.27 (3.31)	7.04 (3.76)	Welch's $F(3, 44.02) = 45.54$, $p < .001$	$\eta^2 = .61$
Evidence Memory						
Set C	7.43 (.68)	6.79 (1.18)	5.77 (1.77)	6.82 (1.44)	$F(3, 80) = 5.67$, $p = .001$	$\eta^2 = .18$
Intrusions	N/A	3.28 (3.36)	5.27 (2.23)	N/A	$t(28.48) = 2.16$, $p < .05$	$\eta^2 = .11$
Source Memory						
Set C	7.19 (1.17)	4.63 (1.71)	5.95 (1.46)	6.05 (1.40)	$F(3, 80) = 10.54$, $p < .001$	$\eta^2 = .28$
Intrusions	N/A	3.00 (1.49)	4.95 (1.56)	4.74 (1.96)	$F(1, 61) = 8.00$, $p = .001$	$\eta^2 = .21$
Confidence Ratings						
Evidence	4.64 (.28)	4.46 (.44)	4.35 (.44)	4.56 (.53)	$F(3, 79) = 1.74$, $p = .165$	
Sources	4.69 (.37)	3.76 (.55)	4.01 (.52)	4.27 (.59)	$F(3, 78) = 11.63$, $p < .001$	$\eta^2 = .31$

Note. Higher scores indicate a greater magnitude of each variable.

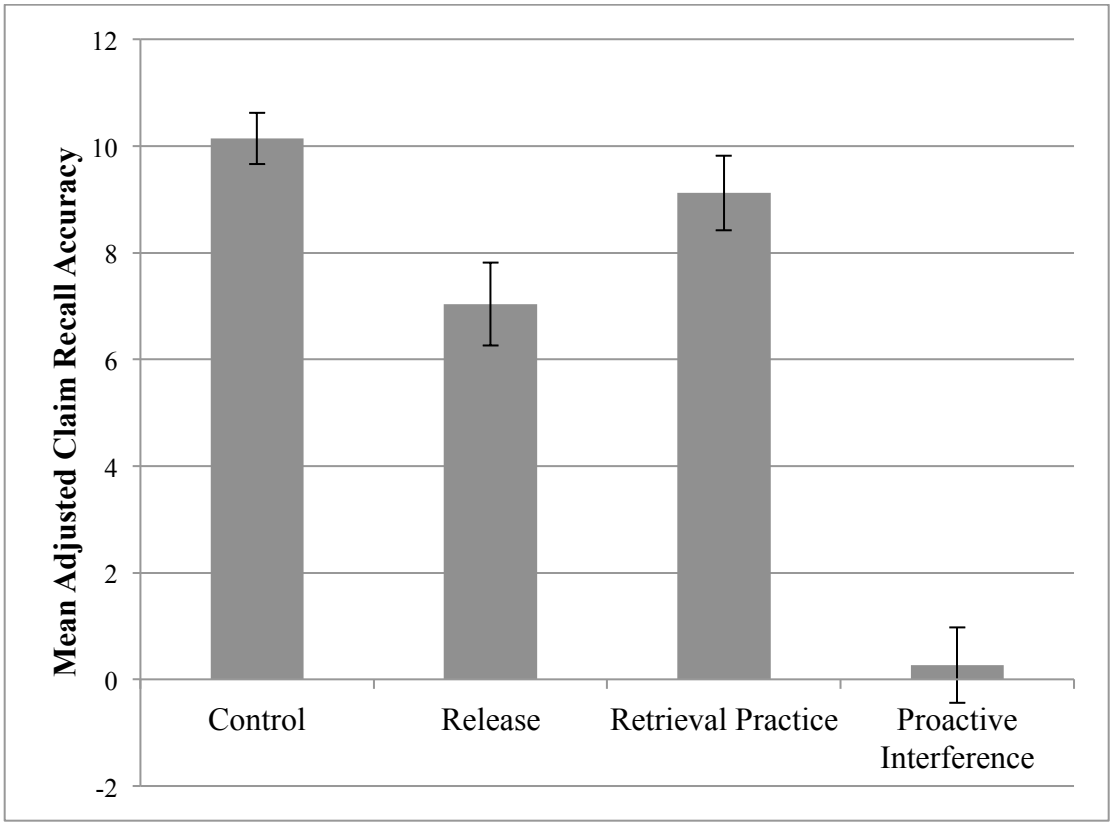

Figure 1. Mean adjusted claim recall accuracy for the target eight texts.

Appendix

Staying Connected via Social Media

Author Occupation: social worker

Publication Venue: *Woman's Day* magazine

Type of Publication: featured column

Social networking sites are beneficial for society because they help senior citizens feel more connected to society. According to a 2010 study, people over 74 years of age are the fastest growing demographic on social media sites with the percentage quadrupling from 2008 to 2010. Seniors report feeling happier due to online contact with family and their church community.

Social Media Promotes Self-diagnosis

Author Occupation: internal medicine specialist

Publication Venue: *American Journal of Public Health*

Type of Publication: journal article

Social networking sites are detrimental for society because they encourage amateur advice and self-diagnosis for health problems, which can lead to harmful or life-threatening results. For example, a Twitter search for "eczema" found in the first 100 results, 84 were spam and several others gave harmful advice.

Social Media Consumes Too Much of Our Time

Author Occupation: online contributor

Publication Venue: Fox News

Type of Publication: opinion/editorial

Social networking sites are detrimental for society because they entice people to waste time. 40% of 8 to 18 year olds spend an hour a day on social media sites. When people receive new tweets or Facebook messages, they take 20-25 minutes on average to return to their original task. In 30% of cases, it took two hours to fully return attention to the original task.

Social Networking Sites Help to get the Music to the Fans

Author Occupation: public relations agent

Publication Venue: www.frontandcenter.com

Type of Publication: blog entry

Social networking sites are beneficial for society because they offer a way for musicians and artists to build audiences even if they don't have a corporate contract. 64% of teenagers listen to music on YouTube, trumping radio and CDs as today's music "hit-maker." For example, pop star Justin Bieber was discovered on YouTube when he was 12 years old, and, in 2012 at 18 years old, his net worth was estimated at $80 million.

Social Networking Can Impact Political Change

Author Occupation: journalist

Publication Venue: National Public Radio (NPR) broadcast

Type of Publication: news brief

Social networking sites are beneficial for society because they can facilitate political change. Social networking sites give social movements a quick, no-cost method to organize, disseminate information, and mobilize people. In 2011, the Egyptian uprising, organized largely via social

media, motivated tens of thousands of protest demonstrations and, ultimately led to the resignation of Egyptian President Mubarak.

Cyberbullying is Rampant on Social Networking Sites

Author Occupation: single parent

Publication Venue: *Chicago Tribune* newspaper

Type of Publication: featured article

Social networking sites are detrimental for society because they facilitate cyberbullying. 49.5% of students reported being the victims of bullying online and 33.7% reported committing bullying behavior online. Adults can also be victims of cyberbullying, from social, familial, or workplace aggression being displayed on social media sites.

Social Media Helps the Workforce

Author Occupation: CEO

Publication Venue: www.monster.com

Type of Publication: advertisement

Social media sites are beneficial for society because they help employers find employees and job-seekers find work. 64% of employment companies use two or more social networks for recruiting because of the wider pool of applicants and more efficient searching capabilities. Moreover, one in six recent hires credit social media in helping them find their current job.

Social Networking Sites Spread False Information

Author Occupation: staff writer

Publication Venue: www.cnn.com/technology

Type of Publication: news report

Social networking sites are detrimental for society because they enable the rapid spread of unreliable and false information resulting in real world impacts. Earlier this year, hackers took over the Associated Press Twitter account and falsely claimed that there had been explosions at the White House and that the president was hurt, triggering financial panic at the stock market (which plunged 143 points).

From: Christopher Wayne Whitehead (cwhitehd) on behalf of Institutional Review Board
Sent: Wednesday, September 02, 2015 3:57 PM
To: Jason Lawrence Braasch (jlbrasch)
Subject: IRB Approval 2941

Hello,
The University of Memphis Institutional Review Board, FWA00006815, has reviewed and approved your submission in accordance with all applicable statuses and regulations as well as ethical principles.
PI NAME: Jason Braasch
CO-PI:
PROJECT TITLE: Comprehending Texts
FACULTY ADVISOR NAME (if applicable):
IRB ID: #2941
APPROVAL DATE: 09/02/2015
EXPIRATION DATE:
LEVEL OF REVIEW: Exempt Modification
Please Note: Modifications do not extend the expiration of the original approval
Approval of this project is given with the following obligations:
1. If this IRB approval has an expiration date, an approved renewal must be in effect to continue the project prior to that date. If approval is not obtained, the human consent form(s) and recruiting material(s) are no longer valid and any research activities involving human subjects must stop.
2. When the project is finished or terminated, a completion form must be completed and sent to the board.
3. No change may be made in the approved protocol without prior board approval, whether the approved protocol was reviewed at the Exempt, Exedited or Full Board level.
4. Exempt approval are considered to have no expiration date and no further review is necessary unless the protocol needs modification.
Approval of this project is given with the following special obligations:

Thank you,
James P. Whelan, Ph.D.
Institutional Review Board Chair
The University of Memphis.
Note: Review outcomes will be communicated to the email address on file. This email should be considered an official communication from the UM IRB. Consent Forms are no longer being stamped as well. Please contact the IRB at IRB@memphis.edu if a letter on IRB letterhead is required.