THE ROLE OF RETRIEVAL DURING STUDY:
FROM SELF-PACED STUDY TIME AND OVERT REHEARSAL

BY

GEOFFREY LOGAN MCKINLEY

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Arts in Psychology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2015

Urbana, Illinois

Advisor:

Professor Aaron S. Benjamin
Abstract

The reminding effect (Tullis, Benjamin, & Ross, 2014) describes the increase in recall for a study word when a related word is presented later in the study list. However, because the process of reminding is thought to occur during study, measures of test performance are only indirect indices of reminding and consequently subject to additional influences. The present research seeks evidence of reminding during encoding, and relates those study behaviors to the more remote consequences at test. In both experiments, participants were presented pairs of related and unrelated words that were separated by various lags. In Experiment 1, self-paced study times were collected and used to index the on-line process of reminding. In Experiment 2, participants were instructed to rehearse out loud anything that came to mind during study. In both experiments, the reminding effect was observed at retrieval and reminding-relevant activities during encoding were predictive of later memory. In Experiment 1, more study time allotted to an associate of an earlier item predicted better memory for its earlier counterpart. Similarly, in Experiment 2, overt rehearsal partially mediated the benefit to memory engendered by semantic associations across items. These two different on-line measures of reminding at study help establish a causal link between the action of reminding at study and later consequences at test.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>EXPERIMENT 1</td>
<td>8</td>
</tr>
<tr>
<td>Method</td>
<td>8</td>
</tr>
<tr>
<td>Results</td>
<td>10</td>
</tr>
<tr>
<td>Discussion</td>
<td>12</td>
</tr>
<tr>
<td>EXPERIMENT 2</td>
<td>14</td>
</tr>
<tr>
<td>Method</td>
<td>14</td>
</tr>
<tr>
<td>Results</td>
<td>15</td>
</tr>
<tr>
<td>Discussion</td>
<td>20</td>
</tr>
<tr>
<td>GENERAL DISCUSSION</td>
<td>22</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>24</td>
</tr>
<tr>
<td>FIGURES AND TABLES</td>
<td>29</td>
</tr>
</tbody>
</table>
**Introduction**

Ongoing events have the potential to remind one of previous episodes. Thinking about relationships between current events and past ones can clarify previous or current ambiguities, aid in generalizing or contrasting across a set of related experiences, and direct one’s attention to relevant events in the future. Indeed, many higher-level cognitive processes would be simply impossible without the ability to consider events that are separated—sometimes distantly so—in time and space (Benjamin & Ross, 2010).

More formally, research on reminding demonstrates how a later episode can elicit retrieval of a prior episode during encoding, and what the consequences of that retrieval are for the involved memories (Benjamin & Ross, 2010; Benjamin & Tullis, 2010; Hintzman 2011; Jacoby & Walhheim, 2013). Events that share some form of similarity seem to effortlessly elicit memories of each other; such similarity-motivated retrieval might be thought to impact category development and provide a basis for generalization. A new event can elicit the retrieval of an older event, which can strengthen memory for the older event (Tullis, Benjamin, & Ross, 2014). Further, reminding can also mediate whether an earlier event interferes with or facilitates the retrieval of a later-formed memory (Wahlheim & Jacoby, 2013). This basic mechanism can subserve a variety of desirable, higher-order cognitive activity such as the re-organization of information, new insight, and flexible knowledge structures.

From a pedagogical perspective, understanding the process of reminding could aid in our understanding of the attributes that a learner extracts during a learning episode. Ultimately, educators could use this information to assess and even influence a student’s comprehension (see Ross, 1984, 1987). Specifically, this information could be used to facilitate memory by
informing educators on how to structure lesson plans in ways that will effectively control the timing, direction, and specificity of a reminding event.

Researchers have asserted a role for reminding in many higher-order cognitive processes, including category learning (Medin & Schaffer, 1978; Brooks, Norman, & Allen, 1991; Ross, Tenpenny & Perkins, 1990), analogical reasoning (Gick & Holyoak, 1980), ambiguity resolution (Ross & Bradshaw, 1994; Tullis, Braverman, Benjamin, & Ross, 2013), early acquisition of skill learning, problem solving, and generalization (Ross, 1984). More recently, researchers have begun to consider the causes and consequences of reminding as they relate to memory. The reminding effect refers to the boost in memory for an item that is followed by a semantic associate, compared to an unrelated item (Tullis, Benjamin, & Ross, 2014). For example, queen is better recalled if king is shown later in the study list than if it is not. Tullis et al. (2014) attributed this enhancement in memory to the covert retrieval elicited by the later, related event. However, because the consequences of reminding are being measured at test, there is some interval between when the initial reminding occurs (during study), and when it’s observed (at test). As a result, these data are limited by the fact that this effect is an indirect index of reminding. The purpose of the present research is to examine the consequences of reminding by collecting measures during encoding, in addition to retrieval. The current experiments use self-paced study time (Experiment 1) and overt rehearsal (Experiment 2) in attempt to observe the immediate consequences of reminding, in conjunction with measures of test performance. Rather than relying exclusively on memory performance as a proxy for the reminding process, this approach permits a direct examination of reminding by revealing any immediate consequences of retrieval during encoding. In doing so, it can offer a more complete understanding of the effects of reminding on memory.
Evidence for reminding

Evidence for the effects of reminding on memory has been found with absolute recency judgments (Hintzman, Summers & Block, 1975; Jacoby & Wahlheim, 2013), relative recency judgments (Tzeng & Cotton, 1980; Winograd & Soloway, 1975), list discrimination (Jacoby, Wahlheim, & Yonelinas, 2013), free recall (Tullis, Braverman, Benjamin, & Ross, 2014; Tullis, Benjamin, & Ross, 2014, Exps, 1a-1c), and cued recall (Jacoby & Wahlheim, 2013; Tullis, Benjamin, & Ross, 2014, Exps. 3a-3b). There is also evidence for a role for reminding in understanding the effects of repetitions on memory (e.g. Benjamin & Tullis, 2010; Greene, 1989; Hintzman, 2004, 2010), though reminding is difficult to unambiguously identify in test performance when the stimulus that is thought to elicit reminding is nominally the same as the one that the learner is being reminded of. That is, because the repetitions of stimuli are identical, subsequent recall of the word does not distinguish between presentations. Nonetheless, the core logic of reminding can be extended to related pairs, providing a framework that unifies research on repetition and semantic associates (Benjamin & Ross, 2010; Benjamin & Tullis, 2010).

Indeed, hints of reminding can also be found in memory research, using related words, even in the absence of any explicit mention of reminding (e.g. Batchelder & Riefer, 1980; Glanzer, 1969; Jacoby, 1974; Robbins & Bray, 1974; Bruce & Weaver, 1973; Rundus, 1971). For example, Jacoby (1974) presented participants one or two members of different semantic categories. Members that shared a common category were separated by varying lags. It was found that when participants were encouraged to verify whether the current word shared a common category with any previous words in the list (the n-back condition), memory performance was enhanced for items that shared a common category. Further, this pattern changed very little across levels of lag. In contrast, when this “looking-back” behavior was
restricted to the item that immediately preceded the current item (the one-back condition),
memory performance dropped substantially for nonzero lags. In light of these findings, it was
concluded that related items can interactively promote memory only if they are ‘brought together
during study’ (i.e. reminding). Similarly, Rundus (1971, Exp. 3) presented a list composed of
category and non-category members to participants, and instructed them to rehearse out loud any
words that came to mind during study. Results showed that participants were much more likely
to rehearse study items that shared the same category as the current item. Importantly, this
pattern did not just reflect maintenance of category members that had been immediately
rehearsed in the previous trial. Specifically, this tendency also reflected a return of members
that had been dropped from the rehearsal protocol in the previous trial. Consistent with theories
of reminding, not only were the number of rehearsals influenced by category membership, but
they also predicted gains in subsequent memory performance.

Recently, Benjamin and Tullis (2010) have suggested that when two items in a list are
related to each other, the presentation of the second item (P2) during study may elicit retrieval of
the first item (P1). A successful retrieval of P1 serves to enhance memory for that item. This
makes a basic prediction that memory for P1 should be enhanced when it’s followed by a related
item, compared to when it’s followed by an unrelated item. This increase in memory is the
reminding effect reported by Tullis et al. (2014). One important alternative to reminding of this
finding is that these results could be due to retrieval effects. In particular during free recall, the
retrieval of one item may elicit retrieval of its related counterpart, or the retrieval of one item
may interfere with the performance of other items in the list (Ratcliff, Clark, & Shiffrin, 1990).
To rule out such test-based explanations, Tullis et al. (2014, Exps. 3A-3B) used probes to
independently cue each item, and also manipulated whether P1 or P2 was tested first. In doing
so, the reminding effect persisted, allowing them to rule out the possibility that the observed
effect was solely due to retrieval processes that operated at test.

These findings emphasize the challenge of separating the contribution of reminding
during study from related factors that may influence memory performance at test. When the
process of interest is theorized to occur during study, test performance is a distant marker of that
process that may be contaminated by events that followed the original reminding event. The
current experiments report immediate evidence of reminding during self-paced study and overt
rehearsal, and relate these behaviors to evidence of the enhancement of memory on a later test
theorized to result from reminding.

On-line measures of reminding

Although measures of memory performance have been much of the focus in research on
reminding, there has been some research that has collected measures during study. For example,
Jacoby and Wahlheim (2013) employed the “looking-back” procedure from Jacoby (1974), in a
similar design. Specifically, participants were presented with related pairs that were separated
by varying lags. As in the prior study, they were asked to verify whether the current item shared
a category with the preceding item (one-back), or any of the previous items in the list (n-back).
At test, participants were shown pairs of words and asked to judge which item had been
presented more recently. In conjunction with recency judgments at test, response times for the
verification task were collected during study. They replicated the results of Jacoby (1974) and,
importantly, found that verification time was longer in the n-back condition than in the one-back
condition. Their procedure differs from Tullis et al. (2014) in that subjects were told to explicitly
“look back” in time, but does indicate that behavior during the study task—in their case, the
category-membership judgment—might be related to later memory.
Wahlheim and Jacoby (2013, Exp. 2) showed that judgments during study that reflect reminding can be used to understand the presence or absence of interference as well. They found that, when a change was detected across two presentations of a paired-associate set (A-B, A-D), the new item was studied for longer than when change was not detected. In addition, the second item did not suffer interference when compared to a baseline (A-B, C-D) condition when the change was detected. In contrast, proactive interference was present when the change was not detected.

Other measures of online processing have been shown to be relevant to reminding. For example, Tullis et al. (2014, Exp. 3B) found that a word was given a higher judgment of learning (JOL) when a related word, compared to an unrelated word, had preceded it. Interestingly, a higher P2 JOL predicted an enhancement in memory for its P1 counterpart but not for the P2 item itself!

Reminding also influences the resolution of ambiguity at the time of reminding. Tullis et al. (2013) reported an experiment in which participants were asked to write sentences using homographic words. Those words were presented on complex background scenes. When a homograph was preceded by a biasing cue that appeared on the same background—conditions that are thought to elicit reminding—it influenced the interpretation of the homograph and also enhanced subsequent memory for the cue. For example, given a presentation of the word “bank”, participants were more likely to interpret it in way that is consistent with “river” if that item had been presented with the same background as “bank”, than if that item had been presented on a different background. Similarly, memory for the two words was enhanced when they shared a common background.
Self-paced study time has proven to be a useful measure in probing on-line processing of materials and relating that processing to eventual memory performance. For example, Shaughnessy, Zimmerman, and Underwood (1972, Exp. 3) found that participants spent less time studying a repeated word if it had been recently studied than if it had been more distantly studied, lending credence to the view that some of the advantage of separating repetitions in time owes to the greater attention people are willing to pay after a longer interval. Self-pacing has the potential to be particularly helpful for understanding reminding because it can be measured without calling undue attention to the relationship between the stimuli (cf. Wahlheim & Jacoby, 2013), thus revealing effects of reminding under conditions with lower demand characteristics.

Overt rehearsal, or “think aloud” measures, also provide a useful way to probe learning (Rundus, 1971). Measures of overt rehearsal have been used to help understand primacy and recency effects, distinctiveness, and the spacing effect (Tan & Ward, 2000). Others have used this general approach to examine word frequency effects in recall (Ward, Woodward, Stevens, & Stinson, 2003), as well as other effects related to output order (Ward, Tan, & Grenfell-Essam, 2010).

In the current experiments, we use these two measures—study time and overt rehearsal—to examine the immediate consequences of reminding and the relationship between those measures and the downstream memory effects that have been taken to reveal reminding. Experiment 1 allowed participants to self-pace their study. Specifically, lists were constructed consisting of pairs in which the P1 was related or unrelated to P2. Additionally, P1 and P2 were separated by a lag of either zero or two intervening items. If a related P2 elicits retrieval of its P1 counterpart, then presumably this would influence how much time is spent studying P2, and enhance memory for P1.
Experiment 1

Method

Participants. Seventy-nine introductory-level psychology students for the University of Illinois at Urbana-Champaign participated in exchange for partial course credit. One participant’s data were dropped for not finishing the experiment within the time allotted for the study.

Materials. 36 primary associate pairs were collected from the University of South Florida Free Association Norms database (Nelson, McEvoy, & Schreiber, 2004). These pairs were picked with the goal of obtaining a moderate associative strength between two words in a related pair (as defined by the database). Similarly, the test cues for each word were chosen to be moderately associated to its intended target, but no more than minimally associated to the word’s related counterpart, nor to any of the other items in the list. Finally, two filler items were selected that were minimally related to any of the cues or targets. Across participants and items, the average forward and backward associative strengths for the related pairs were .53 (SD = .173) and .54 (SD = .161). All of the unrelated pairs had a forward and backward associative strength of zero. The average cue-target forward and backward associative strength was .072 (SD = .057) and .014 (SD = .028), respectively. Across all other cue-target combinations, the average forward and backward associative strengths were near zero.

One list structure was created which contained 24 slots (i.e. 2 positions per slot) and 2 positions for filler items. For each half of the list, there were six slots for the shorter lag (zero intervening items), six slots for the longer lag (two intervening items), and one position for a filler item. Lag was manipulated only to ensure a greater range of conditions under which to detect the reminding effect and is not meaningfully related to any theoretical variables of interest.
here (cf. Tullis, Ross, & Benjamin, 2014). For each participant, the pairs were randomly assigned to either the related or the unrelated condition, with the constraint that each condition be equally represented within each half of the study list. If a pair was assigned to the related condition, the right-hand member always served as P2, and the left-hand member always served as P1 (e.g. salt-pepper). If a pair was assigned to the unrelated condition, an additional pair was randomly selected from the remaining pairs, and the left-hand member of that pair served as P1. Therefore, P2 was controlled for across conditions. This yielded 48 words items plus 2 fillers for the study list. For the test, each pair was randomly assigned to have either P1 be tested first or P2 tested first, with the constraint that each test condition was represented an equal amount of the time within each of the four study conditions (when collapsed across study list halves). The fillers were not tested.

**Design.** The experiment used a 2 (semantic condition) x 2 (lag) within-subjects design. Items within a pair were either related or unrelated to each other. Pairs were separated by a lag of either zero or two intervening items.

**Procedure.** Participants were given the following instructions: “Study each word as long as you need in order to remember it. When you are ready to move on to the next word, press the Space Bar. The list will be long but please do your best to remember these words for a later memory test”. They were further instructed to ask the experimenter for any clarification. Words were presented sequentially in the center of the screen on a white background, in a 40-point black Arial font and remained on the screen until the participant pressed the space bar. Once the participant pressed the space bar, a blank white screen was presented for 500 ms before the next word appeared. After all of the items were presented, participants were given an independent-probe cued recall test. They were instructed the following: “Next you will begin the memory
test. One-at-a-time, you will be given a specific cue that relates to one of the studied words and
the first letter of the target studied word. Please type in the whole word from the study list that
corresponds to the cue and letter combination. For example, you might have studied the word
"dog." The cue and letter combination you may be given at test could be "cat - d--" OR "furry -
d--" or "loyal - d--". The cues and letter combos relate to one specific word from the study list”.
These instructions were presented sequentially with no more than three sentences being
presented before progressing to the next set of instructions. Again, they were instructed to ask
the experimenter for any clarification.

**Results**

**Cued recall performance.** Collapsed across lag, cued recall performance of P1 was higher for
related pairs (M = 0.62) than for unrelated pairs (M = 0.54), t(77) = 3.21, p < .05, d = 0.36, r =
0.59 (see Figure 1).\(^1\) At a lag of zero, cued recall performance of P1 was higher for related pairs
(M = 0.63) than for unrelated pairs (M = 0.54), t(77) = 3.04, p < .05, d = 0.34, r = 0.42 (see
Figure 2). At a lag of two, cued recall performance of P1 was numerically higher for related
pairs (M = 0.60) than for unrelated pairs (M = 0.55); however, this was did not reach
significance, t(77) = 1.67, p > .05, d = 0.19, r = 0.42 (see Figure 3).

Collapsed across lag, cued recall performance of P2 was higher for related pairs (M =
0.55) than for unrelated pairs (M = 0.49), t(77) = 3.10, p < .05, d = 0.35, r = 0.54. At a lag of
zero, cued recall performance of P2 was numerically higher for related pairs (M = 0.56) than for
unrelated pairs (M = 0.51); however, this did not reach significance, t(77) = 1.76, p > .05, d =
0.20, r = 0.41. At a lag of two, cued recall performance of P2 was higher for related pairs (M =
0.54) than for unrelated pairs (M = 0.46), t(77) = 2.56, p < .05, d = 0.29, r = 0.33.

\(^1\) r refers to the within-subjects correlation across groups.
P2 study time. Collapsed across lag, P2 study time was less for related pairs (M = 4.73 seconds), than for unrelated pairs (M = 5.70), t(77) = -3.27, p < .05, d = 0.37, r = 0.93 (see Figure 4). At a lag of zero, P2 study time was less for related pairs (M = 4.02 seconds), than for unrelated pairs (M = 5.57), t(77) = -4.02, p < .05, d = 0.45, r = 0.88 (see Figure 5). At a lag of two, P2 study time was numerically less for related pairs (M = 5.43) than for unrelated pairs (M = 5.83); however, this was not significant, t(77) = -0.91, p > .05, d = 0.10, r = 0.88 (see Figure 6).

If P2 elicits retrieval of its related P1 counterpart, then the time spent during the presentation of P2 should be predictive of memory for P1. To evaluate this, a logistic regression analysis was conducted on each participant’s data for each condition, collapsed across lag. Because there were only 12 observations per fit, 7 participants’ fits were dropped due to perfect separation. That is, for some of the data sets, there existed a study time in which all of the words that were studied for at least this amount of time were recalled, and any words that fell short of this time were not recalled. As a result, a participant’s data were dropped if this occurred in at least one of the two conditions. That is, if a participant’s data from at least one of the two conditions could not be fit, then the all of the participant’s data were dropped from this analysis.

To compare these fits across conditions, the y-intercept and the slope of each fit were combined into a 50% effective rate parameter (Cramer, 1991), which is essentially the x-intercept of the logit function. This measure indicates how much study time is required for a given item in order to recall it at a 50% probability. Therefore, this predicts that the related condition should yield a lower effective rate parameter. These data are shown in Figure 7. To test whether the effective rate was lower for the related than for the unrelated condition, a nonparametric (within-subjects) permutation test was conducted on the medians of each
condition (see Ernst, 2004). The decision to compare the medians instead of the means was made because there were some extreme outliers in the data (see Table 1). The observed median difference was compared to the sampling distribution generated by the data, and revealed that the probability of obtaining this difference under the null is unlikely, *p* < .05. This analysis suggests that when P2 is related to P1, less P2 study time is required to achieve P1 cued recall performance that is comparable to performance of the unrelated condition. One possibility is that these data were observed because P1 was encoded better, which not only potentiated its later retrieval, but also reduced P2 study time. To explore this possibility the entire analysis was repeated using P1 study time as a predictor of P2 recall (see Table 2), however, there was no significant difference between related and unrelated pairs. This suggests that these data were observed primarily because of a process that was initiated at P2 (i.e. the retrieval of P1 at P2), rather than at P1 (e.g. encoding of P1).

**Discussion**

To summarize, three main results, when the data are collapsed across lag, are consistent with the idea that study remindings influence later memory. First, less study time was allocated to P2 when a semantic associate had preceded it. Second, cued recall performance was higher for both P1 and P2 when the two items were related, compared to when they were unrelated. Third, when the two items were related, gains in P1 memory performance were predicted by less P2 study time than when the two items were not related.

Including self-paced study time in conjunction with memory performance allows a more direct examination of reminding than relying solely on cued recall performance. However, a change in study time doesn’t unequivocally index the process of reminding. Consequently, although this measure is collected during the presentation of P2, it’s still somewhat indirect in
that it relies on the inference that less study time indicates retrieval. Another caveat is that self-paced study precludes any experimenter control over presentation time. To alleviate these two concerns, Experiment 2 used overt rehearsal (Ward et al., 2003; Rundus, 1971; Grenfell-Essam, Ward, & Tan, 2013) at study as an index of reminding. In this procedure, study time can be controlled, and reminding can be inferred much more directly from rehearsal protocols.
Experiment 2

Method

Participants. Seventy-one introductory-level psychology students from the University of Illinois at Urbana-Champaign participated in exchange for partial course credit. One subject’s data were dropped due to little familiarity with the English language, as was reported in a post-experiment questionnaire. This was also evident in their overt rehearsal (i.e. strong foreign accent) and their low recall performance (M = 0.17). Another participant did not rehearse any of the words. As a result, their data were only included in the cued recall analyses.

Materials. The words used in this experiment were the same as in Experiment 1. The average forward and backward associative strengths for the related pairs were slightly different (both Ms = .53, SD = .167 and .165, respectively), due to variations in their random assignment to each condition. All of the unrelated pairs had forward and backward associative strengths of zero. Across all conditions, the left-hand member served as P1 and the right-hand member served as P2 with a 50% probability. For the other percentage of the time, the left-hand member served as P2 and the right-hand member served as P1. If a pair was assigned to the related condition, the items in the pair would be presented, separated by either four or six intervening items. If a pair was assigned to the unrelated condition, one of the items served as P1. An additional pair was randomly selected from the remaining pairs, and one the items served as P2. The lags between pairs were extended to a lag of four, and a lag of six. The lags were extended by four in attempt to ensure that the rehearsal of P1 was not influenced by temporal proximity at the shorter lag (Tan & Ward, 2008). Similarly, it was thought that the rehearsal would extend the likelihood of reminding to larger lags, as rehearsal may reduce the functional lag, or the lag
between the last rehearsal of P1 and the presentation of P2. All other aspects of the design were the same as Experiment 1.

**Design.** The experiment used a 2 (semantic condition: related or unrelated) x 2 (lag: four or six intervening items) within-subjects design.

**Procedure.** The procedure was identical to Experiment 1 with a few exceptions. First, words were presented for a duration of 1.5 seconds followed by an interstimulus interval of 4 s for rehearsal. Once the item disappeared, participants were instructed to “first say that word out loud, and then say anything that comes to mind that you think will help you remember the words”. After the instructions had been read, participants had to paraphrase the instructions to the experimenter. Experimenters listened for three main points in the participants’ description: (1) to say that word out loud first, (2) to use the time before the next word to say any words that come to mind, and (3) to speak clearly and loudly into the microphone. If these points were not communicated to the experimenter, or if the participant asked for clarification, the main points were verbally repeated by the instructor.

**Results**

A rehearsal set (RS) for each item was defined by the words that were rehearsed from presentation of a given item to the end of the interstimulus interval that followed it. For each RS, a 1 was given to each word that was present in the RS, irrespective of its rehearsal frequency within the RS, and a 0 if the word was not present in the RS. All the measures of rehearsal were computed based on these sets. Reminding theory hypothesizes that P2 is more likely to remind a learner of P1 when the two words are related, which should result in a boost in memory performance for P1. Predictions of overt rehearsal behavior are based on three measures: P1 rehearsal at P2 presentation, co-rehearsal of P1 and P2, and recency. Based on the first measure,
if P2 reminds that participant of a semantically associated P1, then P1 should be present in the
RS of P2 more often for related pairs than for unrelated pairs. It is possible that the effects of
reminding are not limited to the moment of P2 presentation and test performance, but also
influence rehearsal throughout the list. For example, perhaps the initial reminding event during
P2 presentation creates a contingency between items in related pairs such that rehearsal of one
member of a pair is more likely given rehearsal of its counterpart. If this were the case, then P1
and P2 should be in the same RS more often when they are related than unrelated; that is, they
should be co-rehearsed more often when they are related as opposed to unrelated. Finally, if a
contingency between P1 and P2 is present for related pairs, then it would follow that the
rehearsal of related words would persist longer than unrelated words throughout study. That is,
the last rehearsal of a related word should appear closer to the end of list than an unrelated word.
This measure of recency was computed as a distance measure from the end of the list, which
indicated the difference between the length of the list and the position of the RS that contained
the last rehearsal of an item.

*P1 rehearsal at P2.* Collapsed across lag, P1 was much more likely to be rehearsed
during the presentation of P2 when the two were related (M = 2.41) compared to unrelated (M =
0.41), t(68) = 9.21, p < .05, d = 1.11, r = 0.16 (see Figure 6). This indicates that there was a 40%
probability that P1 would be rehearsed during the presentation of P2 when the two were related,
compared to a 7% chance when they were unrelated.

*Proportion of co-rehearsals.* To examine whether related pairs were more likely to be
co-rehearsed throughout the list, co-rehearsal of a pair was counted as a binary variable where a
1 was given to a pair if it was co-rehearsed at all, and 0 if it was never co-rehearsed. Collapsed
across lag, related pairs were more likely to be co-rehearsed (45%) than unrelated pairs (9%),
t(68) = 9.50, p < .05, d = 1.14, r = 0.16^2 (see Figure 7).

Although this is an informative finding, it only distinguishes between whether P1 and P2 were co-rehearsed or not. That is, it does not indicate how many times these pairs were co-rehearsed. To more closely examine this finding, the number of times in which a pair was co-rehearsed, given that it was co-rehearsed once, was computed. Collapsed across lag, co-rehearsed pairs were co-rehearsed an average of 1.85 times when related, and 1.19 times when they were unrelated, t(69) = 2.70, p < .05, d = 0.32, r = 0.68 (see Figure 8).

Recency. To evaluate the average recency of the last rehearsal, the distance from the end of the list was computed. This was done by subtracting the word number of the last rehearsal from the total number of words in the study list. P1 dropped out of rehearsal later when it had been followed by a related item (M = 21.22), compared to an unrelated item (M = 22.79), t(68) = -3.59, p < .05, d = 0.43, r = 0.71. Although the last rehearsal of P2s were numerically closer to the end of the list when it was part of a related pair (M = 18.63) than when it was not (M = 19.21), the effect was not significant, t(69) = -1.15, p > .05, d = 0.14, r = 0.26 (see Figure 9).

Cued recall performance. Collapsed across lag, cued recall performance of P1 was higher for related pairs (M = 0.63) than for unrelated pairs (M = 0.60); however, this was not significant, t(69) = 1.40, p > .05, d = 0.17, r = 0.30 (see Figure 10). At a lag of 4, cued recall performance of P1 was lower for related pairs (M = 0.57) than for unrelated pairs (M = 0.61); however, this was not significant, t(69) = -1.16, p > .05, d = 0.14, r = 0.20 (see Figure 11). At a lag of 6, cued recall performance of P1 was higher for related pairs (M = 0.69) than for unrelated pairs (M = 0.58); t(69) = 3.11, p < .05, d = 0.37, r = 0.18 (see Figure 12).

---

^2 Statistical analyses were conducted on the number of pairs that were co-rehearsed as opposed to the proportions of pairs that were co-rehearsed.
Collapsed across lag, cued recall performance of P2 did not differ for related pairs (M = 0.61) than for unrelated pairs (M = 0.58); t(69) = 1.32, p > .05, d = 0.16, r = 0.28. At a lag of 4, cued recall performance did not differ for related pairs (M = 0.60) than for unrelated pairs (M = 0.56); t(69) = 0.98, p > .05, d = 0.12, r = 0.08. At a lag of 6, cued recall performance did not differ for related pairs (M = 0.62) than for unrelated pairs (M = 0.59); t(69) = 0.81, p > .05, d = 0.10, r = 0.15.

As mentioned above, lag was manipulated in order to capture a wider set of conditions in which reminding could occur. Specifically, the reasoning was that a reminding is most likely to occur at short lags, and that an increase in lag should reduce its likelihood. However, the fact that this effect was present at a longer lag, but not present at a shorter lag is quite puzzling. Further, the effect was numerically reversed at a lag of four, such that items that were followed by a semantic associate were recalled at a numerically lower rate than items followed by an unrelated item. Although it’s not clear why the reminding effect did not show up at a lag of four, it has been replicated multiple times (in Tullis et al. [2014, 3A-3B], as well as in our Experiment 1); the (unweighted) average effect size in these three experiments is 0.32. Our failure to produce a similar effect with power estimated from that prior work suggests that performance at shorter lags is somehow altered by overt rehearsal.\(^3\)

For present purposes, we pursued a more fine-grained mediation analysis in which the differential rehearsal across conditions is included. To examine whether the increased rehearsal in the related condition predicts cued recall performance, two mediation analyses were conducted. Specifically, it was of interest whether the number of rehearsals that were given to an

---

\(^3\) A reliable pattern was observed in a pilot study using recognition instead of cued recall that was consistent with these data. Taken together, these findings suggest that overt rehearsal alters the encoding process in a way that makes it distinct from ordinary encoding.
item mediated the relationship between condition and cued recall performance (MacKinnon, 2008; MacKinnon, Krull, & Lockwood, 2000). Because there were only 6 pairs per condition, the data were collapsed across lag to double the number of observations. Our primary analysis was done on P1, which yielded 12 observations for each condition of semantic association. Of course, this test still suffers from very low power. To further increase our statistical power, a second mediation analysis was conducted on all of the items, collapsed across P1 and P2. The necessary parameters were estimated using linear mixed models with random intercepts for participants and items, using the lmer4 software package in R.

More specifically, the following equations were used:

(1) \( \logit(P(Y_{ijk})) = \gamma_{jk} + \tau X_{ijk} \)

(2) \( \logit(P(Y_{ijk})) = \gamma_{jk} + \tau' X_{ijk} + \beta Z_{ijk} \)

(3) \( Z_{ijk} = \gamma_{jk} + \alpha X_{ijk} + \varepsilon_{ijk} \)

where \( \gamma_{jk} = \gamma_{00} + u_{j0} + v_{0k} \), \( i \) denotes the trial, \( j \) denotes the participant, and \( k \) denotes the item, \( \gamma_{00} \) denotes the overall intercept, \( u_{j0} \) denotes the deviation of \( j^{th} \) participant’s intercept from the overall intercept, \( v_{0k} \) denotes deviation of the \( k^{th} \) item’s intercept from the overall intercept, \( \varepsilon_{ijk} \) is the error term, \( X \) is the semantic association condition (related = 1, unrelated = 0), \( Z \) is the number of trials in which an item was included in the rehearsal set, and \( Y \) is whether or not the item was recalled at test (1,0). Figure 13 depicts the model. Table 3 shows the parameter estimates. To derive the p-value for \( \alpha \), a model-comparison approach was implemented which compared the model denoted in equation 3 against the model without condition as a predictor.

To test the hypothesis that the number of trials that a P1 was rehearsed mediates the relationship between condition and memory performance, the confidence interval for the indirect
effect was estimated using a bootstrapping procedure. More precisely, an observation was resampled with replacement $N$ times (i.e. 69), at which point the indirect effect was estimated. This process was iterated 5000 times to generate the sampling distribution. To test the indirect effect, the bias-corrected and accelerated (BCa) 95% confidence interval (Efron & Tibshirani, 1993) was computed using the boot package in R. This interval did overlap with zero $[-0.006, 0.068]$, indicating that the number of rehearsal sets that included a P1 did not reliably mediate the relation between condition and cued recall performance. However, this interval did not overlap with zero $[0.003, 0.043]$ when the analysis was done on both P1 and P2, indicating that the number of rehearsal sets that included a given item did reliably mediate the relation between condition and cued recall performance (see Table 3).

**Discussion**

The overt rehearsal measure showed four results. First, P1 was more likely to be rehearsed upon presentation of its P2 counterpart when the two words were related as opposed to unrelated. In addition, the P2 and P1 were more likely to be rehearsed within the same rehearsal set when they were related. Further, when P1 and P2 were co-rehearsed, they continued to be co-rehearsed more often when they were related. Finally, this dependency associated with related words seemed to extend the rehearsal of P1s but not P2s to a later position in the list. One possibility is that P2 primarily serves some sort of scaffolding function by strengthening memory for P1. However, an alternative interpretation is that P1 has more opportunity to be rehearsed throughout the list simply because it is presented earlier on average.

Memory performance was superior for a word when it had been followed by a word that was semantically associated, although this pattern did not hold at a lag of 4. One possible interpretation for this is that related P2 elicited retrieval of its P1 counterpart, but that retrieval
was not laborious enough to potentiate its later retrieval at test (Benjamin, Bjork, & Schwartz, 1998; Gardiner, Craik, & Bleasdale, 1973). That is, because P1 was closer in temporal proximity to P2 at lag of 4, it may have remained in rehearsal for longer, rendering its effective lag to be smaller than its nominal lag. In a similar vein, it’s possible that at a lag of 4, temporal proximity, rather than semantic association, primarily drove the dependency of P1 and P2 in rehearsal. However, this latter interpretation would not explain why P1 was rehearsed less often during P2 presentation when the pairs were unrelated. A final possibility is that overt rehearsal alters the way in which participants encode information at study, in such a way that discourages reminding at shorter lags. Interestingly, this mnemonic advantage did not extend to P2 despite the fact the related pairs were co-rehearsed more often than unrelated pairs. This finding is also consistent with the notion that P2 retroactively benefits P1, but not vice versa.

The indirect effect was relatively small ($\alpha\beta = 0.018$), a fact that is primarily driven by the small effect that overt rehearsal had on memory performance. There is evidence that rote rehearsal has little, if any, influence on memory, especially on recall tasks (Benjamin & Bjork, 2000; Craik & Watkins, 1973), and clearly the overt rehearsal task makes rote rehearsal a very easy strategy to implement and other forms of elaborative rehearsal somewhat more difficult.
General Discussion

For both experiments, memory performance for P1 was higher when it had been followed by a related P2, replicating the reminding effect. This boost in memory is thought to result from the retrieval of P1 in response to a related P2 (Benjamin & Tullis, 2010). Research on reminding has typically relied on test performance as a basis for inference. However, this mnemonic benefit is secondary to the changes in ongoing processing that may be elicited by the reminding event. In the current experiments, more direct evidence for reminding was provided by collecting measures at the time in which the initial reminding event is thought to occur.

Importantly, these online measures predicted gains in memory performance that were consistent with reminding theory. In Experiment 1, P2 study time predicted better P1 memory performance when P2 was a semantic associate. In Experiment 2, overt rehearsal partially mediated the relation between condition and later recall performance.

Using overt rehearsal as a proxy for reminding provided a rich dataset, allowing us to probe the process of reminding while controlling for study time. These data showed that P1 was rehearsed more often during a related P2 presentation, and that this greater amount of rehearsal led to more spaced and later rehearsals—all conditions that would be expected to enhance memory. Memory was enhanced, though not consistently across conditions. This may reflect the weak effect that rote rehearsal is known to have on cued recall (Benjamin & Bjork, 2000; Craik & Watkins, 1973). However, as mentioned above, it’s possible that rehearsal alters the way in which learners study, such that the reminding effect is eliminated or reversed at shorter lags. This finding suggests that overt rehearsal may have limited potential in research on reminding. More importantly, part of the benefit to memory for a related pair of items could be explained by different patterns of rehearsal, a result attested to by the mediation analysis.
Reminding has a pervasive influence on our intellectual lives. It provides a mechanism by which the environment helps play a role in determining what information is relevant and should be retrieved. The environment is dynamic, and as a result, stimulus-guided retrievals foster flexibility in one’s memory and knowledge structures. Similarly, when an event reminds one of an earlier event, this can influence how one currently behaves in response to that event, permitting flexibility in one’s behavior and decision-making.
References


Arnold.


Figures and Tables

Figure 1. Experiment 1: Cued recall performance for P1 and P2, collapsed across lag. Black bars = related, Gray bar = unrelated.
Figure 2. Experiment 1: Cued recall performance for P1 and P2, at a lag of 0. Black bars = related, Gray bar = unrelated.
Figure 3. Experiment 1: Cued recall performance for P1 and P2, at a lag of 2. Black bars = related, Gray bar = unrelated.
Figure 4. Experiment 1: P2 study time, collapsed across lag. Black bars = related, Gray bar = unrelated.
Figure 5. Experiment 1: P2 study time, at a lag of 0. Black bars = related, Gray bar = unrelated.
Figure 6. Experiment 1: P2 study time, at a lag of 2. Black bars = related, Gray bar = unrelated.
Figure 7. Experiment 1: Cued recall of P1 as a function of P2 study time, collapsed across lag.

Black line = related, Gray line = unrelated.
Table 1. Experiment 1: Mean and median y-intercepts, slopes, and 50% effective rates for each condition from the logistic fits using P2 study time to predict P1 cued recall.

<table>
<thead>
<tr>
<th></th>
<th>Related</th>
<th></th>
<th>Unrelated</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Median</td>
<td>Mean</td>
<td>Median</td>
</tr>
<tr>
<td>y-intercept</td>
<td>0.86</td>
<td>0.34</td>
<td>0.39</td>
<td>0.39</td>
</tr>
<tr>
<td>slope</td>
<td>0.18</td>
<td>0.08</td>
<td>-0.69</td>
<td>-0.02</td>
</tr>
<tr>
<td>50% effective rate</td>
<td>-9688287.03</td>
<td>1.30</td>
<td>20.93</td>
<td>2.98</td>
</tr>
</tbody>
</table>
Table 2. Experiment 1: Mean and median y-intercepts, slopes, and 50% effective rates for each condition from the logistic fits using P1 study time to predict P2 cued recall.

<table>
<thead>
<tr>
<th></th>
<th>Related</th>
<th>Unrelated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Median</td>
</tr>
<tr>
<td>y-intercept</td>
<td>0.26</td>
<td>0.14</td>
</tr>
<tr>
<td>slope</td>
<td>0.59</td>
<td>-0.01</td>
</tr>
<tr>
<td>50% effective rate</td>
<td>3556029.84</td>
<td>2.56</td>
</tr>
</tbody>
</table>
Figure 8. Experiment 2: Number of instances in which P1 was rehearsed during the presentation of P2, for each lag. Black bars = related, Gray bar = unrelated.
Figure 9. Experiment 2: Proportion of pairs that were co-rehearsed, for each lag. Black bars = related, Gray bar = unrelated.
Figure 10. Experiment 2: Average frequency of co-rehearsal given that it was co-rehearsed once, for each lag. Black bars = related, Gray bar = unrelated.
Figure 11. Experiment 2: Average distance of the last rehearsal of a word from the end of the list, collapsed across lag. Black bars = related, Gray bar = unrelated.
Figure 12. Experiment 2: Cued recall performance for P1 and P2, collapsed across lag. Black bars = related, Gray bar = unrelated.
Figure 13. Experiment 2: Cued recall performance for P1 and P2, at a lag of 4. Black bars = related, Gray bar = unrelated.
Figure 14. Experiment 2: Cued recall performance for P1 and P2, at a lag of 6. Black bars = related, Gray bar = unrelated.
Figure 15. Mediation model.

Number of Rehearsals ($Z$)

Related vs. Unrelated ($X$) → Cued Recall Performance ($Y$)

Indirect effect: $\alpha \beta$

Total effect: $\tau$
Direct effect: $\tau'$

$\alpha$

$\beta$
Table 3. Mediation analysis with only P1, collapsed across lag.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Parameter</th>
<th>Estimate</th>
<th>SE</th>
<th>Z</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td></td>
<td>0.382</td>
<td>0.2268</td>
<td>1.683</td>
<td>.093</td>
<td></td>
</tr>
<tr>
<td>β</td>
<td></td>
<td>0.067</td>
<td>0.0152</td>
<td>4.371</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Total effect</td>
<td>τ</td>
<td>0.218</td>
<td>0.1149</td>
<td>1.897</td>
<td>.058</td>
<td></td>
</tr>
<tr>
<td>Direct effect</td>
<td>τ'</td>
<td>0.193</td>
<td>0.1160</td>
<td>1.663</td>
<td>.096</td>
<td></td>
</tr>
<tr>
<td>Indirect effect</td>
<td>αβ</td>
<td>0.025</td>
<td>0.0199</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4. Mediation analysis with P1 and P2, collapsed across lag.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Parameter</th>
<th>Estimate</th>
<th>SE</th>
<th>Z</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>α</td>
<td>0.321</td>
<td>0.1318</td>
<td>2.433</td>
<td>.015</td>
<td></td>
</tr>
<tr>
<td></td>
<td>β</td>
<td>0.058</td>
<td>0.0125</td>
<td>4.663</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Total effect</td>
<td>τ</td>
<td>0.233</td>
<td>0.0809</td>
<td>2.886</td>
<td>.004</td>
<td></td>
</tr>
<tr>
<td>Direct effect</td>
<td>τ'</td>
<td>0.216</td>
<td>0.0813</td>
<td>2.655</td>
<td>.008</td>
<td></td>
</tr>
<tr>
<td>Indirect effect</td>
<td>αβ</td>
<td>0.019</td>
<td>0.0099</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>