Journal of Memory and Language 114 (2020) 104128

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

The role of retrieval during study: Evidence of reminding from overt rehearsal

Geoffrey L. McKinley*, Aaron S. Benjamin

University of Illinois at Urbana-Champaign, United States

ARTICLE INFO

Keywords:
Reminding
Overt rehearsal
Retrieval

ABSTRACT

Reminding occurs when a stimulus being studied elicits a spontaneous retrieval of a previously-studied stimulus, an event that is more likely to occur when the two stimuli are related in some way. One consequence of reminding is enhanced recall for a word when followed by a related word later in the study list (i.e., the *reminding effect*; Tullis, Benjamin, & Ross, 2014). However, it is difficult to precisely localize this enhancement to reminding that occurs during study. The present research uses a "think-aloud" protocol and uses measures of overt rehearsal as a direct index of reminding *during encoding*, with the goal of relating these measures to the more distant consequences of reminding at test. In two experiments, participants were presented pairs of related and unrelated words that were separated by various lags and instructed to rehearse out loud anything that came to mind during study. The study phase was followed by a recognition test in Experiment 1 and a cued recall test in Experiment 2. In both experiments, prior related words were more likely to be spontaneously rehearsed during the interval following a related word. In Experiment 2, this pattern of rehearsals was shown to be predictive of later memory, strongly implying a link between the action of reminding at study and memory enhancement at test. Overt rehearsal partially mediated the benefit to memory engendered by semantic associations across items, indicating that reminding is an important route by which semantics exert an effect on episodic memory.

Introduction

Events and experiences can exhibit some similarity to one another. This fact allows one to use past experiences to aid in making sense of the current situation. Thinking about previous relevant experiences can allow one to compare events and make use of instructive structural similarities among temporally distant experiences, or to detect important differences between ostensibly similar events. Such information can be used in developing sophisticated semantic structures, which can be used to predict future events, subsequent planning, and enhance future learning (e.g., Smith, Hasinski, & Sederberg, 2013). Previous experience can also be used to guide ones' behavior in the current moment, permitting one to make on-line inferences about a current event. Such inferences can inform the decision-making process or direct one's attention to relevant characteristics of the current event (see also Wahlheim & Zacks, 2019). This similarity enables consideration of events across time and space. As such, many higher-order cognitive processes are likely dependent on such similarity (Benjamin & Ross, 2010).

Reminding is a theoretical mechanism whereby an ongoing event can elicit retrieval of a previous event (Benjamin & Ross, 2010; Benjamin & Tullis, 2010; Hintzman, 2011; Jacoby & Wahlheim, 2013). Events that are similar make such reminding more likely. When such a retrieval occurs, there are a variety of consequences. Reminding can support higher-order cognitive activities such as the development of flexible knowledge structures, re-organization of information, and new insight. Retrieval of similar events also informs one's immediate situation by guiding one's attention, allowing one to plan more effective strategies, and make more accurate predictions and better choices (Smith et al., 2013; Wahlheim & Zacks, 2019). Empirically, reminding has been a useful mechanism in explaining a variety of different phenomena, including aspects of judgment and decision-making (Gilovich, 1981; Hintzman, Asher, & Stern, 1978), generalization, early acquisition of skill learning and problem solving (Ross, 1984), category learning (Brooks, Norman, & Allen, 1991; Medin & Schaffer, 1978; Ross, Perkins, & Tenpenny, 1990), and ambiguity resolution (Ross & Bradshaw, 1994; Tullis, Braverman, Ross, & Benjamin, 2014). Recently, researchers have begun to revisit the concept of reminding in explaining basic memory phenomena such as the spacing effect (Benjamin & Tullis, 2010; Wahlheim, Maddox, & Jacoby, 2014), memory for related words (Tullis, Benjamin, & Ross, 2014), and proactive interference and facilitation (Jacoby, Wahlheim, & Kelley, 2015; Putnam,

* Corresponding author at: School of Psychology, Georgia Institute of Technology, 654 Cherry St. NW, Atlanta, GA 30332, United States.
E-mail address: Geoffmck59@gmail.com (G.L. McKinley).

https://doi.org/10.1016/j.jml.2020.104128
Received 23 September 2019; Received in revised form 27 April 2020; Accepted 7 May 2020
0749-596X/

Wahlheim, & Jacoby, 2014; Wahlheim & Jacoby, 2013). Evidence for reminding in memory research has also been found using a variety of different measures such as absolute recency judgments (Hintzman, 2010), relative recency judgments (Jacoby & Wahlheim, 2013; Tzeng & Cotton, 1980; Winograd & Soloway, 1985), list discrimination (Jacoby, Wahlheim, & Yonelinas, 2013), spacing judgments (Hintzman, Summers, & Block, 1975), free recall (Hintzman et al., 1978; Tullis, Benjamin et al., 2014, Exps, 1a-1c; Tullis, Braverman, et al., 2014), and cued recall (Jacoby & Wahlheim, 2013; McKinley, Ross, & Benjamin, 2019; Tullis, Benjamin et al., 2014, Exps. 3a-3b).

The *reminding effect* describes the enhancement in memory performance that is observed for an item when it is followed by a semantically-related item during study (Tullis, Braverman, et al., 2014). For example, *senator* (P1) is more likely to be recalled from a study list if *president* (P2) was presented later in the list than if it was not. Tullis, Benjamin et al. (2014) argued that this mnemonic benefit is due to the covert retrieval of P1, elicited during the presentation of P2. It should be noted that the reminding effect describes a finding that is evident only on a later test, even though the reminding event is thought to occur during study. This effect is different from much of the research on higher-order cognitive processes, in which reminding is often assessed in closer proximity to the reminding event. In that literature, the reminding event and the assessment of the consequences of reminding often occur at the same time. However, memory research on reminding is limited by the fact that the action of reminding is theorized to occur during study, even though this particular (mnemonic) consequence of reminding is observed at test. The difficulty is that assessing the effects of reminding on a later test is susceptible to influences that occur between the action of reminding and the measurement of reminding.

An inference that the presentation of a word (P2) elicited retrieval of an earlier semantic associate (P1) requires one to assume that retrieval of P2 *at test* did not remind one of P1. However, this assumption may not be valid (e.g., see Jacoby & Wahlheim, 2013). It is quite possible that when memory for P2 is tested, P1 is remembered before or after the retrieval of P2. Consequently, because P1 and P2 are related, reminding can occur at study *or at test*. Indeed, motivated by this concern, Tullis, Benjamin et al. (2014, Exps. 3A-3B) developed a cued-recall testing procedure that used extra-list cues to independently probe memory for each studied item, with the goal of reducing the influence of reminding during the memory test. However, the validity of this procedure depends on how well the modified memory test was successful in independently probing one member of a studied pair but not the other.

Relying solely on test performance also creates problems in interpreting null effects, thereby hindering an understanding of how reminding behaves across various levels of lag. The reminding theory of Benjamin and Tullis (2010) predicts that as the lag between two events increases, the probability of a reminding event will decrease and the potency, or durability, of a successful reminding event, will increase. Put differently, reminding is thought to be more likely but less potent at shorter intervals. Therefore, a null effect at a shorter lag may be because the reminding event did not yield lasting effects (e.g., see Batchelder & Riefer, 1980), or because no reminding occurred. A null effect at a longer lag may be due to an interval that is too long for any reminding to occur, or to a lag was not long enough for one to gain sufficient potency from a reminding to retain the item until test. Finally, a longer lag could also lead to potent reminding, but only for a subset of items that is too small to overwhelm the influences of other factors. Without independent evidence that an item was retrieved during study, there is not enough information to rule out such possibilities as confounding the standard interpretation of the reminding effect. A related issue with relying solely on test performance is that in the absence of more direct evidence of reminding, one can only infer the presence of reminding in the aggregate. By also measuring this process on-line, one has the opportunity to observe reminding on an item level. As such, one can more precisely examine how reminding affects study behavior, as well as how this study behavior *predicts* later memory performance on an item level.

The purpose of the present research is to study the process of reminding more closely by examining the on-line consequences of reminding at study, as well as the mnemonic consequences of reminding at test. To do this, the current experiments use overt rehearsal during study, in addition to test performance, as an index of study behavior and reminding. As a result, this research permits a more direct examination of reminding at the time of reminding. Importantly, this approach also offers an opportunity to link on-line behavior to later memory performance.

Although the focus in reminding research has been on memory performance, there has been some research that included measures of study behavior. For example, Tullis, Benjamin et al. (2014, Exp. 3B) showed that reminding influences judgments of learning (JOL). Specifically, that study found that higher P2 JOLs were associated with better memory for its earlier, related P1 counterpart, but not with P2 recall. Tullis, Braverman, et al. (2014) found evidence for reminding using an on-line measure of ambiguity resolution. They found that a previously presented word was more likely to influence how participants interpreted a later presented homophone when the two words were presented on the same background, compared to a different background. Words were also more likely to be recalled when the two items shared the same background than when they did not. Wahlheim and Jacoby (2013, Exp. 2) found that proactive interference was less likely when participants detected (and recollected) a change across two presentations of a paired-associate set (A-B, A-D) (see also Jacoby & Wahlheim, 2013; Wahlheim, 2014, Exp. 3; Jacoby et al., 2015; Putnam et al., 2014; Wahlheim & Zacks, 2019, Exp. 2). They also found that a paired associate was studied for longer when a change was detected compared to when no such change was detected (see also Wahlheim, 2014).

Self-paced study time has been used as a way to examine on-line study and its relation to memory on a later test. For example, in studying the spacing effect, Shaughnessy, Zimmerman, and Underwood (1972, Exp. 3) found that massed repetitions were given less study time than spaced repetitions, suggesting that massed repetitions were devalued (Greeno, 1970; Underwood, 1969,1970; Waugh, 1970). Using a self-paced study procedure allows researchers to measure the consequences of reminding without making the relationships between stimuli a focal component of the task, thus reducing the potential for demand characteristics. McKinley et al. (2019) found that participants studied words for less time when they had been preceded by a semantic associate, as compared to an unrelated word. Further, more P2 study time predicted better memory for P1. These results suggest that the presentation of a related P2 elicited retrieval of P1, which reduced P2 study time and enhanced memory for P1.

However, self-paced study time also has limitations. For example, it is not obvious, a priori, whether reminding should be associated with a reduction or an increase in P2 study time. It also precludes any experimenter control over presentation time. Depending on how a participant paces their study, there may be less potential for reminding. For example, a participant may not study P1 long enough for the item to still be remembered by the time P2 has been presented. Similarly, one may not study P2 long enough or deeply enough to be reminded of P1.

Overt rehearsal or "think aloud" measures have been used to help understand a variety of other memory phenomena, including word frequency effects in recall (Ward, Woodward, Stevens, & Stinson, 2003), output order (Grenfell-Essam, Ward, & Tan, 2013), primacy and recency effects (Tan & Ward, 2000), distinctiveness and the spacing effect (Rundus, 1971), as well as reminding (Winograd & Soloway, 1985). For example, Rundus (1971, Exp. 4) presented a list composed of randomized category and non-category members to participants, instructing them to rehearse out loud any words that came to mind during study. Participants often rehearsed words that were semantically related to the currently presented word. Interestingly, category members that had dropped out of rehearsal were also more likely to *return* to the rehearsal set upon presentation of a shared category. More rehearsal was associated with better memory, and consistent with theories of reminding, words that shared a category with other words were recalled more often than words that did not share a common category.

Overt rehearsal allows us to preserve control over the participants' study time while still granting participants typical discretion over encoding strategies. Indeed, Rundus (1971) suggested that overt rehearsal could be used to collect data about how participants naturally process information. If reminding were naturally occurring during study, this would presumably reveal itself in the rehearsal output.

Experiment 1

The goal of Experiment 1 was to assess the effects of relatedness on overt rehearsal during study of words. Participants studied pairs of words that were either semantically related (e.g., *lion – tiger*) or not (e.g., *coral – checkers*), with each member of each pair presented individually and at a certain distance (lag) from its partner. Due to the uncertainty over the ideal lag at which to assess these effects, we used a range of lag conditions, with the hope that one or more of these lags would be within the appropriate window to detect the effect of relatedness on rehearsal. This design called for a large stimulus set in order to fill out the design and have sufficient precision within each cell. Given the long study list, for which a traditional recall test would be very difficult, we chose to use a recognition test. Because recognition does not typically reveal a reminding-based benefit at test (see McKinley et al., 2019; Tullis, Benjamin et al., 2014), Experiment 1 focuses primarily on the contents of overt rehearsal.

Method

Subjects. Ninety introductory-level psychology students from the University of Illinois at Urbana-Champaign participated in exchange for partial course credit. A total of twenty-five participants were dropped from the analyses for the following reasons: sixteen only recited the current word, the protocols of four additional participants were inaudible as a result of loud air conditioning, one was inaudible because of a strong accent, one was inaudible and did not consistently recite the current word first (as instructed), two did not recite the current word first, and one experienced a malfunction with the computer program. These exclusions resulted in a total of sixty-five participants. In this experiment, we did not link any demographic data, such as whether a participant was a native English speaker, to their subject number. Data were collected with the aim of obtaining a sample size that was sufficient to observe an effect size of $d = 0.35$ with 0.8 power, when the data were collapsed across lag. This analysis suggested a sample size of 67 participants.

Materials. Ninety-six 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). Finally, sixteen filler items were selected that were minimally related to any of the previously studied words. Across participants and pairs, the average forward and backward associative strengths for the related pairs were very similar ($Ms = 0.51$, $SD = 0.16$ and 0.17, respectively). The average forward and backward associative strengths for the unrelated pairs were approximately zero ($Ms < 0.001$, $SDs = 0.001$ and 0, respectively). For both related and unrelated conditions, the left-hand member served as P1 and the right-hand member served as P2 with a 50% probability (e.g., salt-pepper). For the other half of the time, the left-hand member served as P2 and the right-hand member served as P1 (e.g., pepper-salt). If a pair was assigned to the related condition, the items in the pair were presented with an interval of zero, three, or seven intervening items between the two members. If a pair was assigned to the unrelated condition, an additional pair (e.g., king-queen or queen-king) was randomly selected from the remaining pairs, and the right-hand member of that pair served as P2 (e.g., salt-queen, pepper-queen, king-pepper, or king-salt). This procedure controlled for the specific identities of P1 and P2 presentations across conditions.

To complete this design, 12 pairs were required to fill each of the six conditions with 8 pairs. One list structure was created that contained 48 slots. Sixteen fillers were placed where it was necessary to meet the demands of the list structure. This yielded a study list composed of 112 words. The study list was split into two halves. Each half contained an equal number of slots for each lag, as well as fillers. For each subject, word pairs were randomly assigned to the related and unrelated conditions with the restriction that there were an equal number of word pairs in each condition.

Design. The experiment used a 2 (Condition: related or unrelated) × 3 (Lag: zero, three, or seven intervening items) within-subjects design.

Procedure. Participants were instructed to "repeat the current word that you see out loud, and use the time in between the words to rehearse any words that come to mind that you think will benefit your performance at test." Participants were also instructed to speak loudly and clearly into a microphone and to restrict their practice to English. The instructions were repeated multiple times. After the instructions were presented, participants were asked to paraphrase the instructions back to the experimenter to make sure they understood the task. 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 repeated by the experimenter. Words were presented singly in the middle of the computer screen and remained on the screen for 1.5 s. This was followed by 4 s of a blank screen.

During the recognition test, single words were presented on the screen, just as during the study session, and participants rated how well they recognized each item on a scale of 1 to 4, with 1 meaning "certain new", and 4 meaning "certain old". Participants rated 192 words during the recognition task, 96 of which had been studied, and 96 of which had not. All studied items were tested. Among the old words, 48 of them were from the related condition and 48 were from the unrelated condition. Among the new words, 48 of them were new pairs that were highly associated to each other but not to any of the words on the study list, and the other 48 were related to each of the words in the unrelated condition. This design allowed us to correct for differences in bias between the conditions.

Results

A rehearsal set (RS) for each item was defined by the words that were rehearsed from the moment of presentation of a given item to the end of the interstimulus interval that followed it. For each RS, a value of 1 was assigned to each word that was present in the RS, irrespective of its rehearsal frequency or position within the RS, and a 0 if the word was not present in the RS. All measures of rehearsal reported here were computed based on these sets. Predictions of overt rehearsal behavior are based on three measures: the probability of P1 rehearsal at P2 presentation, the probability of co-rehearsal of P1 and P2 throughout the list, and the recency of P1 rehearsals.

Reminding theory hypothesizes that P2 is more likely to remind a learner of P1 when the two words are related. As such, P1 should be present in the RS of P2 more often for related pairs than for unrelated pairs. It is also 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

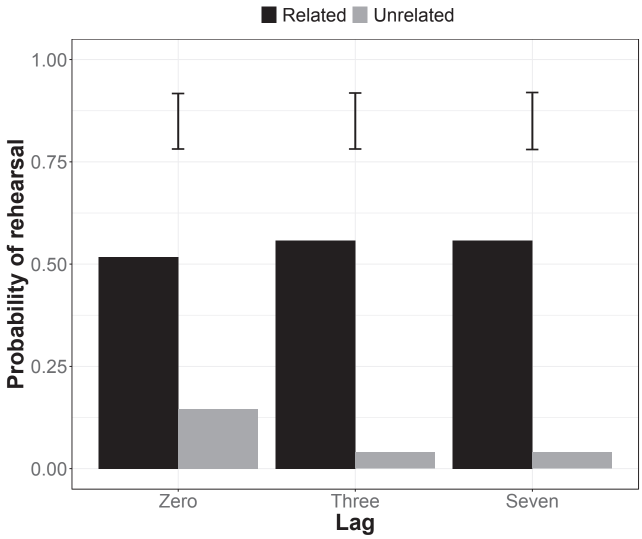

Fig. 1. Experiment 1: Proportion of pairs in which P1 was rehearsed during the presentation of P2 as a function of lag and condition. The error bar represents the 95% CI based on the standard error of the difference between conditions, for each lag separately.

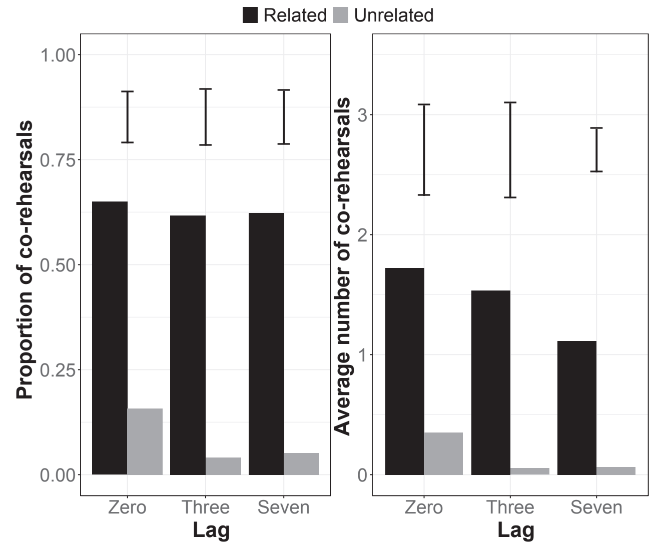

Fig. 2. Experiment 1: Proportion of pairs co-rehearsed as a function of lag and condition (left panel). The average number of co-rehearsals for all pairs (right panel). The error bar represents the 95% CI based on the standard error of the difference between conditions, for each lag separately.

more likely given rehearsal of its counterpart. If this is 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. If a contingency between P1 and P2 is present for related pairs, then it would follow that the rehearsal of related words would foster more prolonged rehearsal throughout the list. 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 last RS that contained a rehearsal of that item.1

P1 rehearsals during P2 presentation. Condition means are shown in Fig. 1. A 2 (Condition: Related, Unrelated) × 3 (Lag: 0, 3, 7) repeated-measures ANOVA was conducted on the frequency with which P1 was rehearsed within the rehearsal set (RS) following its complementary P2 presentation. The main effect of Condition, $F(1,64) = 258.92$, $MSE = 5.29$, $p < .001$, $\eta_G^2 = 0.55$ was significant, with related P1 words being rehearsed more often following presentation of P2 than unrelated P1 words. The effect of Lag was not significant, $F(2,128) = 1.53$, $MSE = 1.94$, $p = .22$, $\eta_G^2 = 0.01$. The Condition × Lag interaction was significant, $F(2,128) = 13.93$, $MSE = 1.06$, $p < .001$, $\eta_G^2 = 0.03$, such that at lag 0, unrelated items were rehearsed more often than at later lags. Collapsed across lag, subjects rehearsed P1 an average of 4.35 (SD = 2.19) times during the P2-rehearsal set when P1 was semantically related, and only 0.61 (SD = 1.06) times when P1 was unrelated.

Number of Co-rehearsals. The data are shown in Fig. 2. A 2 (Condition: Related, Unrelated) × 3 (Lag: 0, 3, 7) repeated-measures ANOVA was conducted on the number of RS within which both P1 and P2 were rehearsed. For each condition, there were 8 possible pairs that could be co-rehearsed. For the sake of clarity, the number of co-rehearsed pairs will be discussed as proportions out of the 8 possible pairs in each condition. The analysis of variance yielded a main effect of Condition, $F(1,64) = 414.35$, $MSE = 4.50$, $p < .001$, $\eta_G^2 = 0.65$, such that related pairs were co-rehearsed more often than unrelated pairs. There was also a main effect of Lag, $F(2,128) = 8.35$, $MSE = 1.68$, $p < .001$, $\eta_G^2 = 0.03$, indicating more co-rehearsal at a lag of zero, as compared to lags three and seven. The Condition × Lag interaction was also significant, $F(2,128) = 4.90$, $MSE = 0.95$, $p < .01$, $\eta_G^2 = 0.01$. The interaction reflects a steady number of co-rehearsals across lag in the related condition, with the number of co-rehearsals dropping dramatically after a lag of zero in the unrelated condition. Averaged across lag, 63% (SD = 0.25) of the related pairs were rehearsed within the same rehearsal block compared to only 8.3% (SD = 0.14) of the unrelated pairs.

Although it is clear that pairs were more likely to be co-rehearsed when they were semantically associated, the above analysis demonstrates only whether a pair was co-rehearsed at least once, or not at all. However, a subject could co-rehearse five related pairs 10 times each, and provide the same score as a subject that co-rehearsed those same pairs only once each. Certainly, the former subject was much more affected by the manipulation of condition, and we sought a measure that would reveal this effect. To do so, an analysis on the frequency of co-rehearsal was conducted. The raw frequency of the co-rehearsals (including zeros) for each subject was averaged for each condition and entered into an ANOVA. This revealed a main effect of Condition, $F(1,64) = 93.93$, $MSE = 1.75$, $p < .001$, $\eta_G^2 = 0.30$, revealing that related pairs were co-rehearsed more often than unrelated pairs. There was also a main effect of Lag, $F(2,128) = 11.96$, $MSE = 0.55$, $p < .001$, $\eta_G^2 = 0.03$, with less co-rehearsal with increasing lag. Finally, there was a Condition × Lag interaction, $F(2,128) = 3.45$, $MSE = 0.47$, $p < .05$, $\eta_G^2 = 0.01$, revealing that the difference in co-rehearsal between related and unrelated pairs increased as a function of lag. Averaged across lag, a given pair was co-rehearsed 1.46 times (SD = 1.39) when the pair was related, compared to 0.16 times (SD = 0.37) when the pair was unrelated.

Recency. Some participants did not rehearse every presented word at least once. Recency was not defined for these trials, and they were dropped for this analysis. In addition, data were excluded from participants who did not rehearse the current word on more than two trials. This led to a loss of data from 12 participants. The data are shown in Fig. 3, and a 2 (Condition: Related, Unrelated) × 3 (Lag: 0, 3, 7) repeated-measures ANOVA was conducted on the average recency of a P1 rehearsal within each condition. A main effect of Condition was observed, $F(1,52) = 19.81$, $MSE = 44.64$, $p < .001$, $\eta_G^2 = 0.06$, in which related P1s were rehearsed closer to the end of the list than unrelated P1s. There was also a main effect of Lag, $F(2,104) = 43.72$, $MSE = 18.61$, $p < .001$, $\eta_G^2 = 0.10$, whereby P1s were rehearsed farther from the end of the list at lag three, and closer to the end of the list at lag seven. The Condition × Lag interaction was not significant, $F(2,104) = 0.81$, $MSE = 44.27$, $p = .44$, $\eta_G^2 = 0.01$. Collapsed across lag, a related-P1′s latest rehearsal occurred 3.33 items later than the

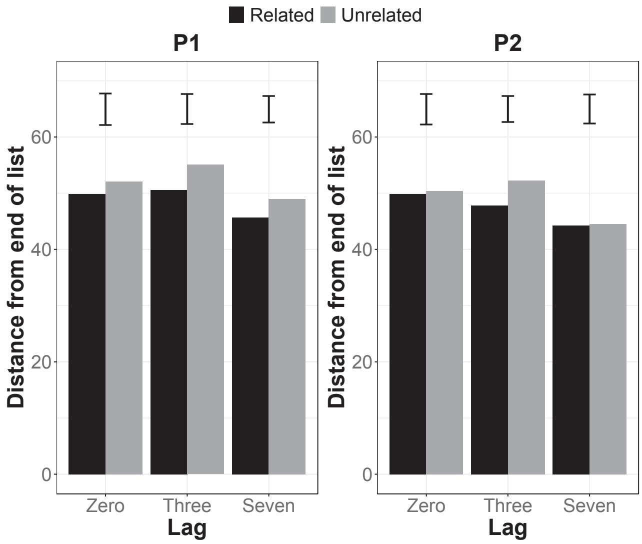

Fig. 3. Experiment 1: Average recency of the last rehearsal for each item type. The error bar represents the 95% CI based on the standard error of the difference between conditions, for each lag separately.

latest rehearsal of an unrelated P1 (on average, 48.71 [SD = 7.24] and 52.04 [SD = 6.89] items away from the last item in the list).

To assess the effect of relatedness on the recency of rehearsal for P2 items, a 2 (Condition: Related, Unrelated) × 3 (Lag: 0, 3, 7) repeated-measures ANOVA was conducted on the average recency of the last rehearsal of P2 within each condition. A main effect of Condition was observed, $F(1,52) = 4.72$, $MSE = 50.70$, $p < .05$, $\eta_G^2 = 0.02$, showing that related P2s were kept in rehearsal for longer than unrelated P2s. There was also a main effect of Lag, $F(2,104) = 62.77$, $MSE = 18.24$, $p < .001$, $\eta_G^2 = 0.16$, with P2s from lag seven being rehearsed closer to the end of the list than P2s from lag three. The Condition × Lag interaction was also significant $F(2,104) = 3.89$, $MSE = 38.07$, $p < .05$, $\eta_G^2 = 0.02$, revealing that the difference in recency was driven by P2s from a lag of three, and not present for lags zero and seven. Collapsed across lag, related-P2′s last rehearsal was 1.74 items after the last rehearsal of unrelated P2 (47.31 [SD = 6.73] and 49.05 [SD = 6.78] items away from the last item in the list).

Recognition. d’ scores were calculated separately for both the related and unrelated condition. False alarms to words that were related to one of the words in the study list were used in calculating d’ for the related condition. False alarms to words that were not related to words from the study list were used to calculate d’ for the unrelated condition. This was done to counteract a well-known bias to say “yes” to words that are related to words on the study list (e.g., Benjamin, 2005; Benjamin & Bawa, 2004) that can create an illusion of a reminding effect. As expected, the hit rate (HR) to related items was higher than to unrelated items (Ms = 0.88 [SD = 0.11] and 0.84 [SD = 0.13]); however, false alarms revealed the same pattern: the average false-alarm rate (FAR) was 0.14 [SD = 0.13] to words from the related condition and 0.10 [SD = 0.13] to words from the unrelated condition. To calculate d’ scores, HR and FAR were corrected via the procedure recommended by Snodgrass and Corwin (1988). Average d’ scores were significantly larger in the unrelated condition (M = 2.56, SD = 0.92) than in the related condition (M = 2.40, SD = 0.73), $t(64) = -2.27$, $p < .05$, $d = 0.28$, $r = 0.77$.[1]

Relation between rehearsal and recognition

If the reminding effect is due to retrieval of a related P1 during P2 presentation, then measures of this reminding action should mediate the relation between semantic association and later recognition of P1. To evaluate this claim, a mediation analysis was conducted (MacKinnon, 2008; MacKinnon, Krull, & Lockwood, 2000) on the data collapsed across lag. A schematic of these analyses is shown in Fig. 4. For all of the models that follow, the necessary parameters were estimated using mixed models with random intercepts for participants and items, using the lme4 software package in R.

More specifically, the following equations were used:

(1) $\text{logit}(P(Y_{ijk})) = \gamma_{jk} + \tau X_{ijk}$
(2) $\text{logit}(P(Y_{ijk})) = \gamma_{jk} + \tau' X_{ijk} + \beta Z_{ijk}$
(3) $\text{logit}(P(Z_{ijk})) = \gamma_{jk} + \alpha X_{ijk}$

where $\gamma_{jk} = \gamma_{00} + u_{j0} + v_{0k}$, i denotes the trial, j denotes the participant, k denotes the item, γ_{00} denotes the overall intercept, u_{j0} denotes the deviation of j^{th} participant’s intercept from the overall intercept, v_{0k} denotes the deviation of the k^{th} item’s intercept from the overall intercept, X is the semantic association condition (related = 1, unrelated = 0), Z denotes whether P1 was present in the rehearsal set corresponding to the presentation of P2, and Y indicates whether or not the item was recognized at test (1,0). In addition, τ denotes the overall effect of condition on memory, τ' denotes the direct effect of condition on memory, α denotes the effect of condition on the rehearsal variable, and β denotes the effect of the rehearsal variable on memory. Table 1 shows the parameter estimates based on the data collapsed across lag.[2]

To test the hypothesis that P1 rehearsal during P2 presentation mediates the relationship between condition and recognition, the confidence interval for the indirect effect was computed. This interval did not overlap with zero [0.19, 0.31], indicating that the P1 rehearsal during P2 significantly mediated the relation between condition and P1 memory performance ($\alpha\beta = 0.25$).[3] This result suggests that changes in the pattern of rehearsals may mediate much of the memory benefit that accrues to P1 as a function of presenting a related stimulus later in the list. However, it should be emphasized that this result can also reflect differences in response bias. That is, related P1s may be more likely than to be endorsed as previously studied words, compared to unrelated P1s, and also rehearsed to a greater degree. We investigate this question more rigorously in Experiment 2.

Discussion

Words followed by a semantic associate were rehearsed more often than words that were not. It is interesting to note that this effect occurred at all lags, even the longest ones. The robustness of this reminding effect stands somewhat in contrast to prior reminding studies in which evidence for reminding (during study) is typically absent by lags of 7 (McKinley et al., 2019). In this study, the presence of overt rehearsal may be affecting the lags for which reminding is present, as it has been shown to affect the shape of the serial position curve in various ways, for example (Fischler, Rundus, & Atkinson, 1970; Tan & Ward, 2008). It may be that overt rehearsal stretches the lag function because rehearsal keeps an item in memory for a longer period of time.

Word pairs that were related were also more likely to be rehearsed *together* than word pairs that were unrelated. In addition, for related

[1] r refers to the within-subjects correlation of scores across conditions.

[2] All of estimates shown are standardized (see MacKinnon & Dwyer, 1993).

[3] The bootstrapping procedure caused many of the iterations to yield models that were unidentifiable. Instead, the confidence interval was computed using the RMediation package in R (see Tofighi & MacKinnon, 2011). This approach constructs the distribution of the product of two random variables (i.e. α and β), based on the estimates of α and β, their standard errors, and the correlation between the two estimates. Changing the value of $\rho_{\alpha\beta}$ within the range of [-0.99, 0.99], did not change the conclusion of the analysis. For simplicity, the parameters reported were estimated assuming no correlation between α and β (i.e., $\rho_{\alpha\beta} = 0$).

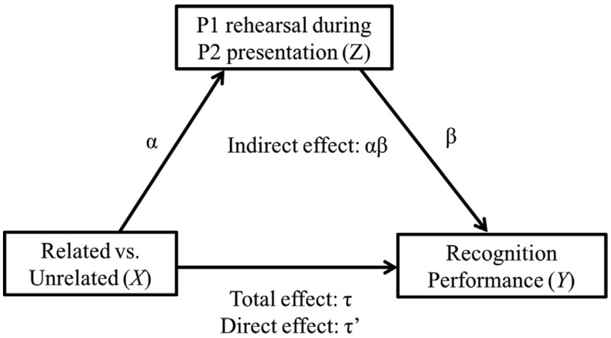

Fig. 4. Mediation model using P1 rehearsal during P2 presentation as a mediator.

Table 1
Mediation analysis with P1 rehearsal at P2 presentation as a mediator of the relation between Condition and P1 hit rate, collapsed across lag. The standard error for the indirect effect was computed assuming no correlation between α and β (i.e., $\rho_{\alpha\beta} = 0$).

Effect	Parameter	Estimate	*SE*	*Z*	*p*
	α	0.635	0.026	**24.58**	< **.001**
	β	0.397	0.046	**8.722**	< **.001**
Total effect	τ	0.143	0.031	**4.633**	< **.001**
Direct effect	τ′	0.002	0.035	0.057	.9545
Indirect effect	αβ	0.252	0.031		

pairs, we computed the frequency of co-rehearsal given that were rehearsed at least once. After excluding two participants who did not contribute to one or more lag conditions, related pairs that were co-rehearsed occurred together in rehearsal 2.38 times on average (SD = 2.40). Finally, both P1 and P2 were rehearsed for a more extended period of time compared to unrelated words.

The Condition × Lag interactions for P1 rehearsal during P2 presentation, co-rehearsals, and P2 recency appear to be partially driven by the fact that at a lag of zero, these measures differed less between conditions than at the longer lags. For example, co-rehearsals were more likely between unrelated words at a lag of zero, and decreased as lag increased. Our interpretation of this finding is that words that were studied in close proximity were more likely to be rehearsed together. These effects reflect the fact that contingencies in overt rehearsal were not solely driven by the semantic similarity of two words, but also their temporal proximity (e.g., Howard & Kahana, 1999).

Although the benefits of reminding were not expected to be seen on a test of recognition, better recognition for unrelated words was certainly unexpected. It seems likely that our attempt to counteract a bias to say "yes" to words related to words from the study list may have made recognition more difficult for the related condition. That is, the lures that were used for the related condition were semantic associates of words from the study list, whereas lures that were used for the unrelated condition were not semantic associates of any of the words from the study list. So it may not be surprising that the unrelated condition exhibited better performance than the related condition.

Intriguingly, P1 rehearsal at P2 presentation mediated the relationship between Condition and performance on the recognition test. However, this result should be interpreted with caution because this analysis does not (and can not) control for potential differences in response bias between conditions. The major focus of Experiment 2 is investigating this relationship in a design that does not have this difficulty.

Experiment 2

In Experiment 2, independent-probe cued recall was used as a measure of test performance. This allows us to examine how reminding affects later memory as well as overt rehearsal. Most importantly, it enables an item analysis of the relationship between overt rehearsal and later memory. These benefits are not without costs, however. As described below, cued recall created a number of new constraints in constructing stimuli. The process of constructing word pairs that met all of these criteria led to a smaller set of stimuli. As a result, we decided to only include two lag conditions instead of three. According to reminding theory (Benjamin & Tullis, 2010), longer lags should reduce the probability of P1 being retrieved during P2 presentation, but should also increase the durability of P1 memory, given that P1 was successfully retrieved. Because rehearsal of related P1 was not substantially affected by lag in Experiment 1, we decided not to include a lag of zero in Experiment 2. In addition, we decided to use a lag of four instead of three for the shorter lag. This was done to reduce P1 rehearsal for unrelated word pairs, with the assumption that participants would have difficulty rehearsing five words in a cumulative forward order (Tan & Ward, 2000,2008). Finally, because our interest was to simply observe a reminding effect at one or both levels of lag, we used a lag of six instead of seven for the longer lag, in order to maximize our chances of observing a reminding effect at test.

Method

Subjects. Seventy-one introductory-level psychology students from the University of Illinois at Urbana-Champaign participated in exchange for partial course credit. A post-experiment questionnaire was given to each participant which asked whether they were native English speakers. One participant did not understand the question due to a lack of proficiency with the English language. As a result, this participant's data were dropped. Data from seven participants were also dropped because they only rehearsed the current word, and data from one participant was dropped because they did not rehearse any of the words. Of the 62 participants included in the rehearsal analyses, 11 were non-native English speakers. Data were collected with the goal of observing an effect size of d = 0.35 with 0.8 power. This analysis suggested a sample size of 67 participants.

Design. The experiment used a 2 (Condition: related or unrelated) × 2 (Lag: four or six intervening items) within-subjects design.

Materials and Procedure. Thirty-six primary associate pairs were collected from the University of South Florida Free Association Norms database (Nelson et al., 2004), in which each pair shared a moderate associative strength. Independent test cues for each word (e.g., *gavel)* were chosen to be moderately associated to its intended target (e.g., *hammer*), but no more than minimally associated to the word's related counterpart (e.g., *nail*), nor to any of the other items in the list. Finally, two filler items were selected to fill out the design. They were chosen to be minimally related to any of the independent cues or study words.

The average forward and backward associative strengths for the related pairs were very similar (both Ms = 0.53, both SDs = 0.17). All of the unrelated pairs had forward and backward associative strengths of zero. The shorter lag between pairs was extended to four in an attempt to ensure that the rehearsal of P1 was not influenced solely by temporal proximity (Tan & Ward, 2000, 2008). The longer lag was six items. All other aspects of the design were the same as Experiment 1.

Results

Rehearsal performance

P1 rehearsal at P2. As shown in Fig. 5, P1 was much more likely to be rehearsed during the presentation of P2 when the two words were related (M = 0.45, SD = 0.28) than when they were unrelated (M = 0.08, SD = 0.10), $t(61) = 9.92$, $p < .001$, $d = 1.26$, $r = 0.06$. This result replicates Experiment 1.

Proportion of co-rehearsals. The data are shown in Fig. 6, where, collapsed across lag, related pairs were more likely to be co-rehearsed

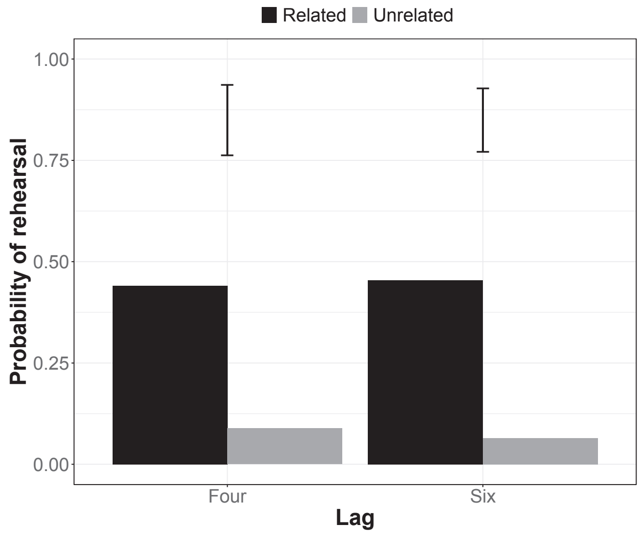

Fig. 5. Experiment 2: Proportion of pairs in which P1 was rehearsed during the presentation of P2, as a function lag and condition. The error bar represents the 95% CI based on the standard error of the difference between conditions, for each lag separately.

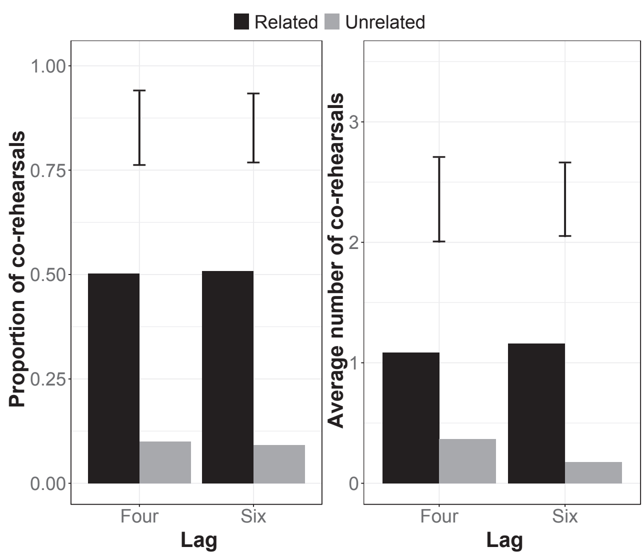

Fig. 6. Experiment 2: Proportion of pairs co-rehearsed as a function of lag and condition (top panel). The average number of co-rehearsals for all pairs (bottom panel). The error bar represents the 95% CI based on the standard error of the difference between conditions, for each lag separately.

(M = 0.51, SD = 0.29) than unrelated pairs (M = 0.10, SD = 0.11), t(61) = 10.67, p < .001, d = 1.35, r = 0.04. It should be noted that it was not uncommon for a given pair to be co-rehearsed later in the list even when P1 was absent from the rehearsal set of P2. That is, because co-rehearsal was not limited to P2 presentation, a pair that is co-rehearsed does not imply that P1 was present in the rehearsal block of P2. Collapsed across lag, pairs were co-rehearsed an average of 1.12 times (SD = 0.98) when they were related and 0.27 times (SD = 0.51) when they were unrelated, t(61) = 7.47, p < .001, d = 0.95, r = 0.41. These effects clearly replicate Experiment 1 and indicate that rehearsal strategies take advantage of semantic patterns in the input.

Recency. As in Experiment 1, participants who did not rehearse a study item at all on more than two trials were dropped from the recency analyses. As a result, data from one additional participant were dropped. Words that were not rehearsed at all were also dropped from the analyses. As shown in Fig. 7, P1 items dropped out of rehearsal later

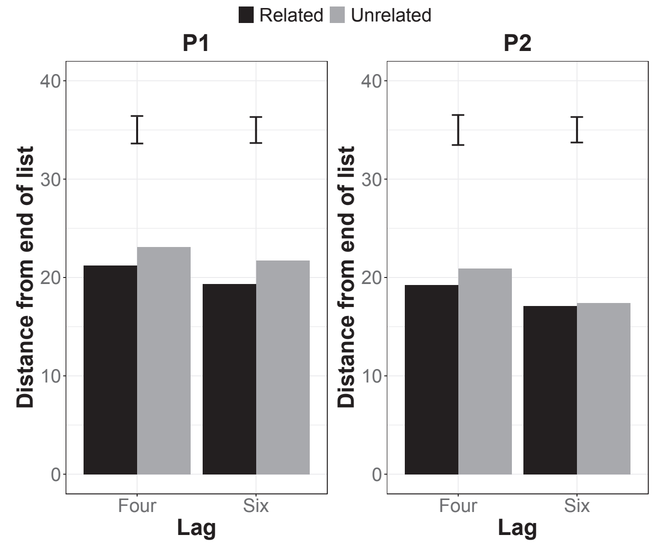

Fig. 7. Experiment 2: Average distance of the last rehearsal of a word from the end of the list, collapsed across lag. The error bar represents the 95% CI based on the standard error of the difference between conditions, for each lag separately.

when it had been followed by a related item (M = 20.28, SD = 4.70) than an unrelated item (M = 22.40, SD = 4.45), t(60) = -4.72, p < .001, d = 0.60, 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 = 18.19, SD = 3.85) than when it was not (M = 19.14, SD = 3.04), the effect was not significant, t(60) = -1.79, p > .05, d = 0.23, r = 0.29.

Cued recall performance

With cued recall testing, a reminding effect is expected on the test (cf. Tullis, Braverman, et al., 2014). These data are shown in Fig. 8. Collapsed across lag, cued recall performance of P1 was higher for related pairs (M = 0.64, SD = 0.16) than for unrelated pairs (M = 0.59, SD = 0.17); however, this effect was not significant, t(61) = 1.71, p > .05, d = 0.22, r = 0.31. At a lag of 4, cued recall performance of

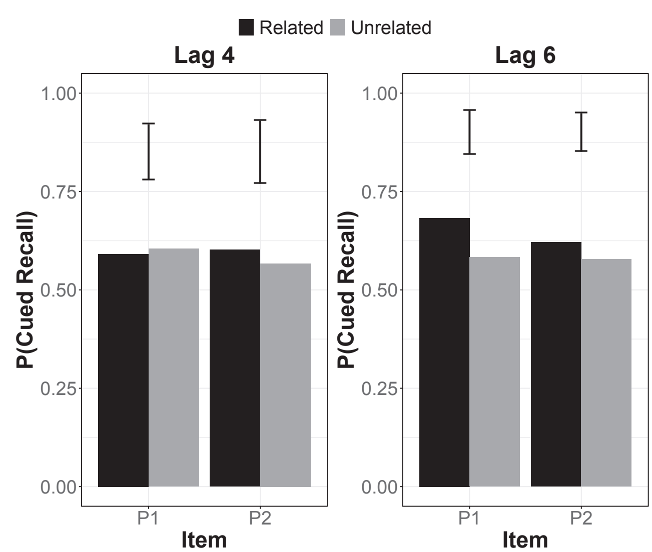

Fig. 8. Experiment 2: Cued recall performance for P1 and P2, at a lag of 4 (left panel) and at a lag of 6 (right panel). The error bar represents the 95% CI based on the standard error of the difference between conditions, for each lag separately.

P1 for related pairs (M = 0.59, SD = 0.22) was not significantly different than cued recall for unrelated pairs (M = 0.60, SD = 0.22; $t(61) = -0.38$, $p > .05$, $d = 0.05$, $r = 0.19$). At a lag of 6, however, cued recall performance of P1 was higher for related pairs (M = 0.68, SD = 0.25) than for unrelated pairs (M = 0.58, SD = 0.23); $t(61) = 2.88$, $p < .01$, $d = 0.37$, $r = 0.18$. As mentioned above, there was some uncertainty in specifying the appropriate size of the lag variable. It seems that a lag of 4 was too short to yield benefits to memory for related P1 in this design. Memory for P2 did not differ across conditions.

It is not clear why the reminding effect did not appear at a lag of four in this experiment. The reminding effect appears to be robust; it has been replicated many times with no known file-drawer effect. (In McKinley et al. [2019, Exp. 2] and Tullis, Braverman, et al., 2014, 3A-3B], the unweighted average effect size is $d = 0.33$.) Our failure to produce a similar effect suggests that the inclusion of overt rehearsal in this design might affect the relationship between lag and memory performance, though such a conclusion is provisional and requires data beyond what these experiments can provide to be evaluated.

Relationship between rehearsal and cued recall

The results from Experiment 2 indicate that overt rehearsal reveals patterns consistent with the operation of reminding. All of the rehearsal patterns replicated Experiment 1. In addition, a reminding effect on memory was evident at test, but only at the longer lag. To test whether P1 rehearsal during P2 presentation mediates the relationship between semantic association and later recall of P1, two mediation analyses were conducted. These analyses were done in the same fashion as Experiment 1, except that the dependent variable was P1 recall, rather than hit rate (see Fig. 4). This test is more definitive, because differences in response bias between relatedness conditions is unlikely to cloud performance on our independent-probe cued-recall test. The first analysis was conducted on the data that were collapsed across lag, which yielded 12 observations for each condition per subject. A second mediation analysis was conducted on only the data from lag 6, which reduced the number of observations by half.[4] A mediation analysis on the data from lag 4 was not pursued because there was no observed effect on memory to mediate. Table 2 shows the parameter estimates derived from the 3 equations above, based on the data that were collapsed across lag. Table 3 shows the parameter estimates based on the data exclusively from lag six.

To test the hypothesis that P1 rehearsal during P2 presentation mediates the relationship between condition and memory performance, the confidence interval for the indirect effect was computed.[5] Collapsed across lag, this interval did not overlap with zero [0.002, 0.097], indicating that the P1 rehearsal during P2 significantly mediated the relation between condition and P1 memory performance. However, it is noteworthy that the effect was quite small, ($\alpha\beta = 0.048$). At a lag of 6, the indirect effect was almost twice the size ($\alpha\beta = 0.100$); its confidence interval also did not overlap with zero [0.032, 0.173]. This combination of results yields provisional evidence that changes in the pattern of rehearsals mediate some of the memory benefit that accrues to P1 as a function of presenting a related stimulus later in the list.

Table 2
Mediation analysis with P1 rehearsal at P2 presentation as a mediator of the relation between condition and P1 memory, collapsed across lag. The standard error for the indirect effect was computed assuming no correlation between α and β (i.e., $\rho_{\alpha\beta} = 0$).

Effect	Parameter	Estimate	*SE*	*Z*	*p*
	α	0.585	0.039	**15.043**	**< .001**
	β	0.081	0.039	**2.080**	**.0375**
Total effect	τ	**0.067**	**0.033**	**2.010**	**.0444**
Direct effect	τ'	0.034	0.037	0.935	.3500
Indirect effect	$\alpha\beta$	0.048	0.023		

Table 3
Mediation analysis with P1 rehearsal at P2 presentation as a mediator of the relation between condition and P1 memory, for lag 6. The standard error for the indirect effect was computed assuming no correlation between α and β (i.e., $\rho_{\alpha\beta} = 0$).

Effect	Parameter	Estimate	*SE*	*Z*	*p*
	α	0.610	0.057	**10.631**	**< .001**
	β	0.164	0.057	**2.903**	**.0037**
Total effect	τ	0.132	0.046	**2.853**	**.0043**
Direct effect	τ'	0.066	0.052	1.289	.1975
Indirect effect	$\alpha\beta$	0.100	0.036		

Discussion

Experiment 2 replicated all four major results evident in Experiment 1 and provided a novel one that helps explain the origin of the memory benefit that results from reminding. First, P1 was more likely to be rehearsed upon presentation of its P2 counterpart when the two words were related. Second, P2 and P1 were more likely to be rehearsed within the same rehearsal set when they were related. Third, when P1 and P2 were co-rehearsed, they continued to be co-rehearsed more often when they were related than when they were unrelated. Fourth, this dependency associated with related words extended the rehearsal of P1s to a later point in the list.

The new result in this experiment is that rehearsal of P1 immediately following P2 presentation was predictive of the degree to which memory for P1 was enhanced, at least at longer lags. This result suggests that the benefits of reminding owe in part to the patterns of rehearsal that relatedness among stimuli foster.

One puzzling finding in this experiment is that, though memory was superior for a word when it had been followed by a word that was semantically associated, this pattern did not obtain at a lag of 4. Previous studies have typically shown this effect at lags in that range (McKinley et al., 2019; Tullis, Benjamin, et al., 2014; Tullis, Braverman, et al., 2014). It is worth considering whether the addition of overt rehearsal changes the function between lag and memory. It is possible, for example, that P2 elicited retrieval of its P1 counterpart, but that the retrieval was not laborious enough to enhance memory (Benjamin, Bjork, & Schwartz, 1998; Gardiner, Craik, & Bleasdale, 1973). That is, because items remain in the active rehearsal buffer for longer during overt than covert rehearsal, P1 was functionally closer to P2 at a nominal lag of 4 than it has been in previous experiments. It is worth remembering that the shorter lag was chosen to be four in anticipation of such an effect: words are rehearsed in a cumulative forward order for up to three words back before a sharp decline in rehearsal is observed (Tan & Ward, 2008). Although the lags in the current experiment were chosen with the intention of eliminating this shift in buffer size, the additional semantic association may have extended the limit. For whatever reasons, it is possible that overt rehearsal alters the way in which participants encode information at study in such a way that obscures the usual benefit of reminding at shorter lags.

[4] In order to maximize the chances of convergence, the BOBYQ optimizer was used for both mediation analyses.

[5] The confidence interval was computed using the RMediation package in R, because the bootstrapping procedure caused many of the iterations to yield models that were nearly unidentifiable. Changing the value of $\rho_{\alpha\beta}$ within the range of [-0.99, 0.99], did not change the conclusion of the analysis. For simplicity, the parameters reported were estimated assuming no correlation between α and β (i.e., $\rho_{\alpha\beta} = 0$).

General discussion

In Experiment 2, recall for P1 was higher when it had been followed by a related P2, replicating the reminding effect (Tullis, Benjamin, et al., 2014; Tullis, Braverman, et al., 2014). This boost in memory is thought to result from the retrieval of P1 in response to a related P2 (Benjamin & Ross, 2010; Benjamin & Tullis, 2010). Research on reminding has typically relied on test performance as a basis for inference. However, this mnemonic benefit is likely secondary to the changes in ongoing processing that are elicited by the reminding event. In the current experiments, more direct evidence for the presence and consequences of reminding was provided by collecting overt rehearsal measures at the time in which the reminding event is thought to occur. Both experiments provided strong evidence in the content of overt rehearsal for the presence of reminding during learning; results from Experiment 2 additionally demonstrated that these operational indices of reminding partially mediated the relation between semantic relatedness and later recall performance, at least when the two words were separated by six interleaved items.

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 greater rehearsal contingencies and was more durable across the study list—all conditions that would be expected to enhance memory. Memory was enhanced, though not consistently across conditions (cf. Benjamin & Bjork, 2000; Craik & Watkins, 1973).

Our choice of cued recall to measure the mnemonic effects of reminding was motivated by the desire to control the order in which items were tested, in order to reduce the confounding influence of reminding at test. Independent-probe cued recall serves as a conservative test of reminding. It is quite likely that a larger reminding effect would have been observed using free recall (see Tullis, Braverman, et al., 2014), in which reminding-induced retrieval can influence memory during study and during the test.

Results from the current experiments suggest that the consequences of reminding are not limited to test performance, but can also influence immediate behavior that is observed during the reminding event. An event that bears a meaningful relationship with a past event is more likely to elicit retrieval of that past event, and the contents of that retrieval influence the ongoing processing of the current stimulus. The retrieval and postprocessing that ensue also enhance memory for the retrieved information. These claims are consistent with the notion that retrieval is a pervasive component of the learning process. Although one function of retrieval is to search for and select stored information, research has shown that this conceptualization is incomplete. For example, research has indicated that retrieval subserves metacognitive monitoring (Benjamin & Bjork, 1996; Benjamin et al., 1998; Tullis, Benjamin, et al., 2014, Exp. 3B) and ongoing learning (Finn & Metcalfe, 2008; Tullis, Fiechter, & Benjamin, 2018; Tullis, Finley, & Benjamin, 2013). Tullis, Braverman et al. (2014) showed that reminding can influence the interpretation of ambiguous events and that that resolution also has consequences for memory.

We have assumed that retrieval underlies P1 rehearsal at P2 presentation, which increases the probability of retrieving P1 at test. This characterization suggests that retrieval underlies rehearsal, reminding, and recall. Indeed, some have argued that rehearsal and recall are driven by the same process (Laming, 2006; Tan & Ward, 2008), and a similar argument could be extended to reminding. However, although retrieval is likely a common underlying process, we think of reminding as driven mostly by the stimulus itself, whereas rehearsal and recall are motivated by higher-order goals and decisions. And, of course, reminding also likely involves substantial ongoing processing, such as comparison, integration, or contrastive processing, when more complex stimuli are used.

Research on the testing effect also demonstrates that retrieval is a potent tool for promoting retention (Benjamin & Pashler, 2015; Karpicke & Roediger, 2007,2008; Tullis et al., 2013), which is another result that blurs the boundaries between different phases of learning. Whereas the testing effect demonstrates that retrieval can improve retention and serve a function that is typically associated with encoding; the reminding effect demonstrates that retrieval can also occur during study and enhance memory via mechanisms similar to the testing effect. These phenomena are both reminders that our cognitive approach to the world does not distinguish between moments in which information is encoded and other moments in which it is accessed, but rather that both are interwoven in most complex cognitive tasks. It is for this reason that concepts like retrieval mode (e.g., Lepage, Ghaffer, Nyberg, & Tulving, 2000)—in which the memory system is accessing rather than encoding information—are misleading. And it is worth remembering this fact when considering the function of teaching—which should not simply be about transmitting new information, but also as an opportunity to remind students of previously learned information. That is, lesson plans and stimuli materials can be structured in ways that encourage retrieval of relevant information, while discouraging information that may interfere with future learning (Ross, 1984,1987).

One concept that is central to theories of reminding is that of *recursive reminding* (e.g., Hintzman, 2004, 2011). Recursion here refers to the idea that the memory traces formed during reminding contain within them the original memory—that is, the information that one is reminded of. An interesting set of results that follows from this idea reveals that subjective reports about detecting and recollecting a change across similar events separated in time predicts downstream consequences on memory (e.g., Jacoby & Wahlheim, 2013, Jacoby et al., 2015; Putnam et al., 2014; Wahlheim & Jacoby, 2013; Wahlheim & Zacks, 2019). When people can detect and report a change, people have better memory for both the original event and the event that cued reminding of that original event. However, if the experience of reminding is not integrated with the reminded event—as revealed by a failure to report a change that had been detected—then the two memories exert interference on one another. Similarly, the reminding effect (Tullis, Braverman, et al., 2014) proposes that, when an event cues reminding of a prior event during study, memory for the original event is enhanced. In the prior work listed above, participants are typically encouraged to pay attention to the relationship between the current item and previous items, leading to an additional benefit in memory for the reminding (i.e., P2) event.

Limitations of the present work

<u>Inconsistency across rehearsal and memory</u>. Reminding theory describes a trade-off between the probability of reminding, and the potency, or durability, of the reminding memory (Benjamin & Tullis, 2010). As the lag increases between two related words, the probability of reminding decreases as the potency of reminding increases. In Experiment 2, reminding at study was apparent in patterns of overt rehearsal at both lags 4 and 6. Yet the boost to memory that is thought to result from reminding was only apparent after a lag of 6. This combination of results was unexpected and inconsistent with our interpretation that rehearsal mediates the effects of relatedness on memory. It is possible that reminding occurred at lag 4, but it was not difficult enough to yield a measurable effect on memory. Of course, this interpretation is entirely post-hoc and its validity cannot be assessed with the current experiments.

<u>Confounding effects of the overt rehearsal procedure on reminding.</u> In the current study, the instructions for rehearsal emphasized that participants were not restricted to rehearsing words from the study list. The wording for these instructions was designed to discourage controlled efforts to seek out reminding events. Nonetheless, there is no way of asking people to tell you what they are thinking about without revealing to them your empirical interest in what they are thinking about, and it is possible that some of the patterns of rehearsal we have

seen owe to demand characteristics inherent in the procedure. Our studies suffer no more or less from this concern than any other work using overt rehearsal, or think-aloud protocols in general. However, one cannot conclude that rehearsing words from the study list necessarily means that participants were motivated to actively search for related events. Rather, it could simply be that participants rehearsed previously studied words because they thought that such a strategy would help them remember words from the study list. In fact, there is evidence that spontaneous reminding occurs even when the task instructions directly emphasize the importance of looking back (Jacoby et al., 2015).

Reminding has an extensive 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 using memory and knowledge in service of current problems. When an event spontaneously reminds one of an earlier event, this can helpfully influence how one currently behaves in response to that event.

Author note

Geoffrey L. McKinley and Aaron S. Benjamin, Department of Psychology, University of Illinois at Urbana-Champaign.

This research was conducted for the first author's master's thesis at the University of Illinois at Urbana-Champaign.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jml.2020.104128.

References

Batchelder, W. H., & Riefer, D. M. (1980). Separation of storage and retrieval factors in free recall of clusterable pairs. *Psychological Review, 87*, 375–397.

Benjamin, A. S. (2005). Recognition memory and introspective remember/know judgments: Evidence for the influence of distractor plausibility on "remember" and caution about purportedly nonparametric measures. *Memory & Cognition, 33*, 261–269.

Benjamin, A. S., & Bawa, S. (2004). Distractor plausibility and criterion placement in recognition memory. *Journal of Memory & Language, 51*, 159–172.

Benjamin, A. S., & Bjork, R. A. (1996). Retrieval fluency as a metacognitive index. In L. Reder (Ed.). *Implicit Memory and Metacognition*. Mahwah, NJ: Erlbaum.

Benjamin, A. S., & Bjork, R. A. (2000). On the relationship between recognition speed and accuracy for words rehearsed via rote versus elaborative rehearsal. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 638–648.

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General, 127*, 55–68.

Benjamin, A. S., & Pashler, H. (2015). The value of standardized testing: A perspective from Cognitive Psychology. *Policy Insights from the Behavioral and Brain Sciences, 2*(1), 13–23.

Benjamin, A. S., & Ross, B. H. (2010). The causes and consequences of reminding. In A. S. Benjamin (Ed.). *Successful remembering and successful forgetting: A Festschrift in honor of Robert A*. Bjork. New York, NY: Psychology Press.

Benjamin, A. S., & Tullis, J. G. (2010). What makes distributed practice effective? *Cognitive Psychology, 61*(3), 228–247.

Brooks, L. R., Norman, G. R., & Allen, S. W. (1991). Role of specific similarity in a medical diagnostic task. *Journal of Experimental Psychology: General, 120*, 278–287.

Craik, F. I. M., & Watkins, M. J. (1973). The role of rehearsal in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 12*, 599–607.

Finn, B., & Metcalfe, J. (2008). Judgments of learning are influenced by memory for past test. *Journal of Memory and Language, 58*, 19–34.

Fischler, I., Rundus, D., & Atkinson, R. C. (1970). Effects of overt rehearsal procedures on free. *Psychonomic Science, 19*, 249–250.

Gardiner, J. M., Craik, F. I. M., & Bleasdale, F. A. (1973). Retrieval difficulty and subsequent recall. *Memory & Cognition, 1*, 213–216.

Gilovich, T. (1981). Seeing the past in the present: The effect of associations to familiar events on judgments and decisions. *Journal of Personality and Social Psychology, 40*, 797–808.

Greeno, J. G. (1970). Conservation of information-processing capacity in paired-associate memorizing. *Journal of Verbal Learning and Verbal Behavior, 9*, 581–586.

Grenfell-Essam, R., Ward, G., & Tan, L. (2013). The role of rehearsal on the output order of immediate free recall of short and long lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 317–347.

Hintzman, D. L. (2004). Judgment of frequency vs. recognition confidence: Repetition and recursive reminding. *Memory & Cognition, 32*, 336–350.

Hintzman, D. L. (2010). How does repetition affect memory? Evidence from judgments of recency. *Memory & Cognition, 38*, 102–115.

Hintzman, D. L. (2011). Research Strategy in the Study of Memory: Fads, Fallacies, and the Search for the "Coordinates of Truth". *Perspectives on Psychological Science, 6*(3), 253–271.

Hintzman, D. L., Asher, S. J., & Stern, L. D. (1978). Incidental retrieval and memory for coincidences. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.). *Practical aspects of memory*. London: Academic Press.

Hintzman, D. L., Summers, J. J., & Block, R. A. (1975). Spacing judgments as an index of study-phase retrieval. *Journal of Experimental Psychology: Human Learning and Memory, 1*, 31–40.

Howard, M. W., & Kahana, M. J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 923–941.

Jacoby, L. L., & Wahlheim, C. N. (2013). On the importance of looking back: The role of recursive remindings in recency judgments and cued recall. *Memory & Cognition, 41*, 625–637.

Jacoby, L. L., Wahlheim, C. N., & Kelley, C. (2015). Memory consequences of looking back to notice change: Retroactive and proactive facilitation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*, 1282–1297.

Jacoby, L. L., Wahlheim, C. N., & Yonelinas, A. P. (2013). The role of detection and recollection of change in list discrimination. *Memory & Cognition, 41*, 638–649.

Karpicke, J. D., & Roediger, H. L. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language, 57*, 151–162.

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science, 319*, 966–968.

Laming, D. (2006). Predicting free recalls. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 1146–1163.

Lepage, M., Ghaffer, O., Nyberg, L., & Tulving, E. (2000). Prefrontal cortex and episodic memory retrieval mode. *Proceedings of the National Academy of Sciences of United States of America, 97*, 506–511.

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. Mahwah, NJ: Lawrence Erlbaum Associates Inc.

MacKinnon, D. P., & Dwyer, J. H. (1993). Estimating mediated effects in prevention studies. *Evaluation Review, 17*(2), 144–158.

MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding, and suppression effect. *Prevention Science, 1*, 173–181.

McKinley, G. L., Ross, B. H., & Benjamin, A. S. (2019). The role of retrieval during study: Evidence of reminding from self-paced study. *Memory & Cognition, 47*(5), 877–892.

Medin, D. L., & Schaffer, M. M. (1978). A context theory of classification learning. *Psychological Review, 85*, 207–238.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida word association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers, 36*, 402–407.

Putnam, A. L., Wahlheim, C. N., & Jacoby, L. L. (2014). Memory for flip-flopping: Detection and recollection of political contradictions. *Memory & Cognition, 42*, 1198–1210.

Ross, B. H. (1984). Remindings and their effects in learning a cognitive skill. *Cognitive Psychology, 16*, 371–416.

Ross, B. H. (1987). This is like that: The use of earlier problems and separation of similarity effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 629–639.

Ross, B. H., & Bradshaw, G. L. (1994). Encoding effects of remindings. *Memory & Cognition, 22*, 591–605.

Ross, B. H., Perkins, S. J., & Tenpenny, P. L. (1990). Reminding-based category learning. *Cognitive Psychology, 22*, 460–492.

Rundus, D. (1971). Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology, 89*, 63–77.

Shaughnessy, J. J., Zimmerman, J., & Underwood, B. J. (1972). Further evidence on the MP-DP effect in free recall learning. *Journal of Verbal Learning and Verbal Behavior, 11*, 1–12.

Smith, T. A., Hasinski, A. E., & Sederberg, P. B. (2013). The context repetition effect: Predicted events are remembered bettered, even when they don't happen. *Journal of Experimental Psychology: General, 142*, 1298–1308.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117*, 34–50.

Tan, L., & Ward, G. (2000). A recency-based account of the primacy effect in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1589–1626.

Tan, L., & Ward, G. (2008). Rehearsal in Immediate Serial Recall. *Psychonomic Bulletin & Review, 15*, 535–542.

Tofighi, D., & MacKinnon, D. P. (2011). RMediation: An R package for mediation analysis confidence interval. *Behavior Research Methods, 43*, 692–700.

Tullis, J. G., Benjamin, A. S., & Ross, B. H. (2014). The reminding effect: Presentation of associates enhances memory for related words in a list. *Journal of Experimental Psychology: General, 143*(4), 1526–1540.

Tullis, J. G., Braverman, M., Ross, B. H., & Benjamin, A. S. (2014). Remindings influence the interpretation of ambiguous stimuli. *Psychonomic Bulletin & Review, 21*, 107–113.

Tullis, J. G., Fiechter, J. L., & Benjamin, A. S. (2018). The efficacy of learners' testing choices. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*, 540–552.

Tullis, J. G., Finley, J. R., & Benjamin, A. S. (2013). Metacognition of the testing effect: Guiding learners to predict the benefits of retrieval. *Memory & Cognition, 41*, 429–442.

Tzeng, O. L., & Cotton, B. (1980). A study-phase retrieval model of temporal coding.

Journal of Experimental Psychology: Human Learning and Memory, 6, 705–716.
Underwood, B. J. (1969). Some correlates of item repetition in free-recall learning. *Journal of Verbal Learning and Verbal Behavior, 8*, 83–94.
Underwood, B. J. (1970). A breakdown of total-time law in free-recall learning. *Journal of Verbal Learning and Verbal Behavior, 9*, 573–580.
Wahlheim, C. N. (2014). Proactive effects of memory in young and older adults: The role of change recollection. *Memory & Cognition, 42*, 1198–1210.
Wahlheim, C. N., & Jacoby, L. L. (2013). Remembering change: The critical role of recursive remindings in proactive effects of memory. *Memory & Cognition, 41*, 1–15.
Wahlheim, C. N., Maddox, G. B., & Jacoby, L. L. (2014). The role of reminding in the effects of spaced repetitions on cued recall: Sufficient but not necessary. *Journal of Experimental Psychology: General, 40*, 94–105.
Wahlheim, C. N., & Zacks, J. M. (2019). Memory guides the processing of event changes for older and younger adults. *Journal of Experimental Psychology: General, 148*, 30–50.
Ward, G., Woodward, G., Stevens, A., & Stinson, C. (2003). Using overt rehearsals to explain word frequency effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 186–210.
Waugh, N. C. (1970). On the effective duration of a repeated word. *Journal of Verbal Learning and Verbal Behavior, 9*, 587–595.
Winograd, E., & Soloway, R. M. (1985). Reminding as a basis for temporal judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 262–271.