



The role of retrieval during study: Evidence of reminding from self-paced study time

Geoffrey L. McKinley^{1,2} · Brian H. Ross^{1,3} · Aaron S. Benjamin¹

© The Psychonomic Society, Inc. 2019

Abstract

The reminding effect (Tullis, Benjamin, & Ross, *Journal of Experimental Psychology: General*, 143[4], 1526–1540, 2014) describes the increase in recall of a studied 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 indirect indicators of the process of reminding and are subject to influences that arise during testing. The present research seeks evidence of reminding *during encoding*. In two experiments, self-paced study times were used to index the online process of reminding. In Experiment 1, pairs of repeated words, related words, and unrelated words were included in a study list. Study times were shorter for words related to prior words in the list, but only when the lag between those two words was short. Relatedness affected study time by inspiring a reduction in the threshold for termination of study for related words under massed conditions. Experiment 2 replicated the reduction in study time for related words and further showed that the study time allotted to an associate of an earlier item predicted better memory for that earlier word on a cued-recall test. In this experiment, an advantage in memory was observed for related words, and self-paced study time of one word during encoding was predictive of later memory for a related word. These results suggest a link between the action of reminding at study, as indexed by changes in the distribution of study time, and later benefits to remembering, as revealed by the reminding effect.

Keywords Reminding · Self-paced study · Study time

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).

At an empirical level, 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 & Wahlheim, 2013). Events that share some form of similarity seem to effortlessly elicit memories of each other; such similarity-motivated retrieval might be thought to affect 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 important higher-order cognitive activities such as the reorganization of information, new insight, and the development of 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). For example, research on analogical transfer and problem solving suggests that such stimulus-guided retrievals are driven

✉ Geoffrey L. McKinley
Geoffmck59@gmail.com

¹ University of Illinois at Urbana-Champaign, Champaign, IL, USA
² School of Psychology, Georgia Institute of Technology, 648 Cherry St. NW, Atlanta, GA 30313, USA
³ Minerva Schools at KGI, 1145 Market St, San Francisco, CA 94103, USA

primarily by the superficial similarity that a current situation or problem shares with the reminded stimulus (Gick & Holyoak, 1980; Ross, 1987). Such information could be used to inform the design of lesson plans that effectively control the timing and direction of reminding events, and perhaps even their specificity.

Researchers have asserted a role for reminding in many higher-order cognitive processes, including category learning (Brooks, Norman, & Allen, 1991; Medin & Schaffer, 1978; Ross, Tenpenny, & Perkins, 1990), judgment and decision-making (Gilovich, 1981; Hintzman, Asher, & Stern, 1978), analogical reasoning (Gick & Holyoak, 1980), ambiguity resolution (Ross & Bradshaw, 1994; Tullis, Braverman, Benjamin, & Ross, 2014), 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 with an unrelated item (Tullis et al., 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 the enhancement in memory to the covert retrieval elicited by the later, related study 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 is observed (at test). As a result, these data are limited by the fact that this effect is an indirect index of reminding. The current experiments use self-paced study time 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 more 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 spacing judgments (Hintzman, Summers, & Block, 1975), absolute recency judgments (Hintzman, 2010), relative recency judgments (Jacoby & Wahlheim, 2013; Tzeng & Cotton, 1980; Winograd & Soloway, 1985), list discrimination (Jacoby, Wahlheim, & Yonelinas, 2013), free recall (Hintzman et al., 1978; Tullis et al., 2014, Experiments 1a–c; Tullis et al., 2014), and cued recall (Jacoby & Wahlheim, 2013; Tullis et al., 2014, Experiments 3a–b). There is also evidence of 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 repetitions of a stimulus are (by definition) identical, subsequent recall of the word does not distinguish which presentation is being remembered. Nonetheless, the core logic of reminding can be extended from related materials to repeated items, 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 consideration of reminding (e.g., Batchelder & Riefer, 1980; Bruce & Weaver, 1973; Glanzer, 1969; Jacoby, 1974; Robbins & Bray, 1974; Rundus, 1971). For example, Jacoby (1974) presented participants one or two members of several sets of 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 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. He concluded that “bringing related items together during study” (i.e., reminding) promotes interaction between items and strengthens memory (p. 495).

Benjamin and Tullis (2010) 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 P1. The prediction follows that memory for P1 should be enhanced when it is followed by a related item, compared with when it is followed by an unrelated item. This effect occurs in tests of recall and has been called the *reminding effect* (Tullis et al., 2014).

One concern in reminding experiments is that the effects of relatedness are difficult to pin down to the encoding phase of an experiment. In particular, during a free recall test, the *retrieval* of one item may elicit retrieval of its related counterpart. To rule out such test-based explanations as the basis for the reminding effect, Tullis et al. (2014, Experiments 3a–b) used unstudied probes (i.e., extralist cues) to independently cue each studied item. The reminding effect persisted, allowing them to rule out the possibility that the observed effect was due to processes that operated exclusively during the 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. For example, covert retrieval of P1 may precede overt

recall of P2 at test, which would incorrectly give the appearance of reminding having occurred during study. Even when the experiment is deliberately designed to probe only one member of a pair, as were Experiments 3a–b of Tullis et al. (2014), it is impossible to control whether the other member was thought about on the test prior to that test trial. The current experiments report online evidence of reminding during self-paced study (Experiments 1 and 2) and relate these behaviors to enhancement of memory (Experiment 2).

Online 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). 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. Consistent with Jacoby (1974), they found that verification time was longer in the n-back condition than in the one-back condition. Though they did not assess the relationship between the study time for an individual item and its memory, this result suggests that behavior during the study task—in their case, the category-membership judgment—might be related to later memory.

Wahlheim and Jacoby (2013, Experiment 2) showed that judgments during study that reflect reminding can be used to understand the presence or absence of interference as well (see also Hintzman et al., 1978). 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 a change was not detected. In addition, the second item did not suffer interference compared with a baseline condition (A-B, C-D) when the change was detected. In contrast, proactive interference was ample 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, Experiment 3b) found that a word was given a higher judgment of learning (JOL) when a related 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. Similarly, Fraundorf, Watson, and Benjamin (2015) found that speakers in a communicative task tended to use less prosodic prominence when describing the second encounter with an item, and that this reduction was related to superior memory for that item.

Reminding also influences the resolution of ambiguity at the time of reminding. Tullis et al., (2014) 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 were designed to elicit reminding—it influenced the interpretation of the homograph and also enhanced subsequent memory for the cue. For example, participants were more likely to interpret the word *bank* in a way that is consistent with *river* if *river* had previously been presented on the same background as *bank* than if it had been presented on a different background. In addition, 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 online processing of materials and relating that processing to eventual memory performance. For example, Shaughnessy, Zimmerman, and Underwood (1972, Experiment 3) found that people 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 (Greeno, 1970; Underwood, 1969, 1970; Waugh, 1970). 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 fewer demand characteristics. Natural self-pacing provides an opportunity to examine reminding when it is not motivated by instructions and cognitive control.

In the current experiments, we use study time to examine the immediate consequences of reminding. In both experiments, participants were allowed to self-pace their study. The goal of the first experiment was to characterize changes in study time under conditions in which we think reminding is likely to occur. The goal of the second experiment was to relate study-time behavior to the downstream memory effects.

Though reminding theory suggests that the presentation of a related P2 can elicit a retrieval of P1, it is not clear whether this act will increase or decrease P2 study time. On the one hand, adding a cognitive act to the study of P2 could increase the overall time devoted to P2. On the other hand, retrieval of P1 may constrain or facilitate the processing of P2 and result in less P2 study time (cf., Fraundorf et al., 2015). This finding would also be consistent with results from the repetition literature, in which the second presentation of a massed pair elicits less self-paced study time than a new item (e.g., Shaughnessy et al., 1972). Priming effects are also consistent with this reduction in study time, though they provide no reason why later memory for P1 would be enhanced (Tullis et al., 2014), nor why such memory effects would be heavily context dependent (Tullis et al., 2014). There have been priming-based

explanations of related phenomena, like the spacing effect (Challis, 1993), but that theory has little to say about memory for related words, or about the contextual dependence seen in studies of reminding.

Experiment 1 was conducted to carefully evaluate whether reminding affects self-paced study time. Consequently, a recognition test was used to measure memory, which allowed a large number of observations to be collected during study. Although there was little expectation that the reminding effect (i.e., the memory benefit) would be revealed by a recognition test (see Tullis et al., 2014), this choice of procedure allowed us to collect a sufficient number of study-time trials to fit theoretical distributions to each individual participant's data. Pairs of repeated words were included as an additional baseline, allowing us to compare the distribution of self-paced study times for related words with both unrelated and repeated control items. In Experiment 2, the relationship between self-paced study time and memory was examined using a cued-recall test, which typically does show the reminding effect (Tullis et al., 2014).

In Experiment 1, participants self-paced their study of words that were repetitions of previous presentations, words that had been preceded by a semantic associate, and words that had been preceded by an unrelated word. According to reminding theory (Benjamin & Ross, 2010; Tullis et al., 2014), the probability of spontaneously retrieving P1 decreases with the degree of semantic association between P1 and P2. That is, P1 retrieval is more likely following a repetition than following a related P2, and more likely following a related P2 than following an unrelated P2. As mentioned above, it is not clear whether such retrieval would lead to longer or shorter study times. The direction of study-time differences between related and unrelated pairs are revealing of the process by which reminding affects cognition. If we see a decrease in study time for related items, then it would suggest that the reduction in study time known to accompany repetition (Shaughnessy et al., 1972) reflects a similar process, and that a common mechanism may underlie the effects of repetition and of semantic relation on memory (cf. Benjamin & Tullis, 2010). Alternatively, an increase in study time for related pairs would suggest that these two situations likely do not share a common mechanism, and that the benefits to memory from relatedness owe to the additional, probably integrative, processing elicited at the time of P2 presentation.

Experiment 1

Method

Subjects Seventy introductory-level psychology students from the University of Illinois at Urbana-Champaign participated in exchange for partial course credit. Six students did not finish the study in time and were dropped, resulting in a total of 64

participants. Our choice of sample size was motivated by an attempt to achieve a statistical power of 0.8 to detect an effect size of $d = 0.35$ for the study-time data.

Materials Ninety-six primary associate pairs were collected from the University of South Florida Free Association Norms (Nelson, McElvov, & Schreiber, 2004). Associated pairs were bidirectionally highly related (mean associative strength = 0.52, $SD = 0.159$) and included synonyms (dinner, supper), antonyms (good, bad), male/female counterparts (king, queen), noun/action pairs (volcano, erupt) and thematically related words (salt, pepper). A list of unrelated words was selected from the same database and was composed of low-frequency items only to limit reminding for those control items. Each control (unrelated) word was assigned to a given pair and always preceded the same P2 when the pair was assigned to be in the unrelated condition. Filler items were also compiled that were matched with the other items for number of orthographic neighbors and frequency. These words were cross-checked against the Free Association Norm database to ensure that none were associates of the previously selected words. Every participant saw the same fillers. One list structure was created that contained 96 slots. Each half contained an equal number of slots assigned to spaced and massed presentations. Fillers were placed where it was necessary to meet the demands of the list structure and were equally distributed between the two halves. This yielded a study list composed of 218 presentations.

For each participant, word pairs were randomly assigned to the related, unrelated, and repetition conditions, with the restriction that there were an equal number of word pairs in each condition. If a pair was assigned to the related condition, a semantic associate preceded P2 in the list. If a pair was assigned to be unrelated, P2 was preceded by an unrelated word. If a pair was assigned to be a repetition, P1 was repeated. Therefore, P2 was the same across the related and unrelated conditions across participants. This was done to control for P2 characteristics so that any differences in study time could be attributed to the identity and location of P1. Note that P2 was not the same item in the repetition condition as in the other conditions. The assignment of conditions to lag was random with the constraint that there was an equal number of each condition within each half of the study list. Descriptive statistics for word frequency and length are reported in Table 1.

A recognition test was constructed that included all of the old items except the fillers. There were an additional 128 new words. This resulted in 160 old and 128 new words. The order of the items on the test was randomized and participants were given unlimited time to complete the test, provided that they did not exceed the time limit of 50 minutes given to complete the experiment. Because item characteristics of P1 and of test lures are (intentionally) confounded with condition, only hit rates for P2 were computed and analyzed.

Table 1 Mean word length and frequency for P1 and P2 for each Lag \times Condition combination, for Experiment 1 and 2

	P1 length	P1 frequency	P2 length	P2 frequency
Experiment 1				
Massed				
Related	5.19 (1.47)	3.20 (0.73)	5.33 (1.60)	3.10 (0.76)
Unrelated	6.21 (1.41)	2.06 (0.53)	5.36 (1.65)	3.08 (0.80)
Repeated			5.14 (1.47)	3.19 (0.70)
Spaced				
Related	5.18 (1.55)	3.20 (0.74)	5.39 (1.71)	3.14 (0.81)
Unrelated	6.11 (1.40)	2.09 (0.54)	5.36 (1.60)	3.12 (0.77)
Repeated			5.18 (1.52)	3.21 (0.74)
Experiment 2				
Massed				
Related	5.56 (1.52)	3.24 (0.70)	5.28 (1.32)	3.34 (0.74)
Unrelated	5.58 (1.61)	3.25 (0.70)	5.51 (1.59)	3.43 (0.67)
Spaced				
Related	5.63 (1.67)	3.22 (0.72)	5.37 (1.51)	3.42 (0.69)
Unrelated	5.54 (1.60)	3.25 (0.73)	5.49 (1.42)	3.41 (0.75)

Note. Standard deviations are in parentheses. Item characteristics for repetitions are reported in the P2 columns, but this label is arbitrary

Design The experiment used a 3 (semantic condition) \times 2 (lag) within-subjects design. The three semantic conditions included repetitions, related pairs, and unrelated pairs. Lag between word pairs was either massed (zero intervening items) or spaced (three intervening items). Lag was manipulated 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. To anticipate, it does, however, raise concern over the opportunity for multiple comparisons and subsequent loss of control over Type I error. We address this concern by replicating key findings in Experiment 2.

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. Some words will be repeated in the list in order to help you remember them. The list will be long, but please do your best to remember as many words as possible.” Words were presented singly in the middle of a white computer screen in 40-point black Arial font and remained on the screen until the participant pressed the space bar. After the participant pressed the space bar, a blank white screen was presented for 500 ms before the next word appeared.

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, where 1 indicated *I am certain I have not seen that word*, 2 indicated *I think I have not seen that word*, 3 indicated *I think I have seen that word*, and 4 indicated *I am certain I have seen*

that word. Participants rated 288 words during the recognition task, 160 of which had been studied and 128 of which had not. All studied items were tested. Among the old words, 64 of them were from the related condition, 64 were from the unrelated condition, and 32 were from the repetition condition.

Results

Study time P2 study times are shown in Fig. 1. Collapsed across lag, P2 study times were shorter when P2 was a repetition ($M = 1.97$ seconds) compared with when P2 was related to P1 ($M = 2.90$ seconds), $t(63) = -4.90$, $p < .001$, $d = 0.61$, $r = .79$ ¹, or when it was unrelated to P1 ($M = 3.23$), $t(63) = -5.71$, $p < .001$, $d = 0.71$, $r = .78$. This was true for both massed, $t(63) = -4.08$, $p < .001$, $d = 0.51$, $r = .59$; $t(63) = -5.13$, $p < .001$, $d = 0.64$, $r = .57$, and spaced, $t(63) = -4.01$, $p < .001$, $d = 0.50$, $r = .77$; $t(63) = -2.98$, $p = .004$, $d = 0.37$, $r = .62$, presentations.

Related P2s were studied for less time than unrelated P2s, $t(63) = -3.08$, $p = .003$, $d = 0.38$, $r = .95$, when collapsed across lag. However, the effect was apparent during massed, $t(6) = -3.94$, $p < .001$, $d = 0.49$, $r = .91$, but not spaced, $t(63) = 0.41$, $p = .686$, $d = 0.05$, $r = .89$, presentations. These results are consistent with the notion that similar processes are occurring during P2 for repetitions and related words, but that the process reducing study time is either less likely or less extensive for related than for repeated words. P1 study time was not analyzed because P1 item characteristics are confounded with condition.²

Recognition performance In analyzing the recognition data, ratings equal to or greater than 3 were considered “yes” responses, and ratings equal to or less than 2 were considered “no” responses. Collapsed across lag, hit rates for related P2s ($M = 0.69$) and unrelated P2s ($M = 0.67$) did not differ, $t(63) = 1.17$, $p = .246$, $d = 0.15$, $r = .84$. Related P2s were recognized at the same rate as unrelated P2s when the pairs were massed ($M_{\text{related}} = 0.70$, $SD_{\text{related}} = 0.20$; $M_{\text{unrelated}} = 0.67$, $SD_{\text{unrelated}} = 0.20$), $t(63) = 1.09$, $p = .280$, $d = 0.14$, $r = .74$, as well as when the pairs were spaced ($M_{\text{related}} = 0.67$, $SD_{\text{related}} = 0.21$; $M_{\text{unrelated}} = 0.66$, $SD_{\text{unrelated}} = 0.21$), $t(63) = 0.71$, $p = .480$, $d = 0.09$, $r = .77$. These findings replicate the result of Tullis et al. (2014) demonstrating no benefit of relatedness on recognition. Unsurprisingly, hit rates for repeated words were numerically much higher when the two presentations were spaced ($M = 0.80$, $SD = 0.19$) than when they were massed ($M = 0.74$, $SD = 0.19$), $t(63) = 2.92$, $p = .005$, $d = 0.37$, $r = .67$. No comparisons with the other conditions were conducted because hit rates in the repetition condition could reflect memory for P1, P2, or both.

¹ Note that r refers to the within-subjects correlation of scores across conditions.

² See Appendix for descriptive statistics for P1 study time.

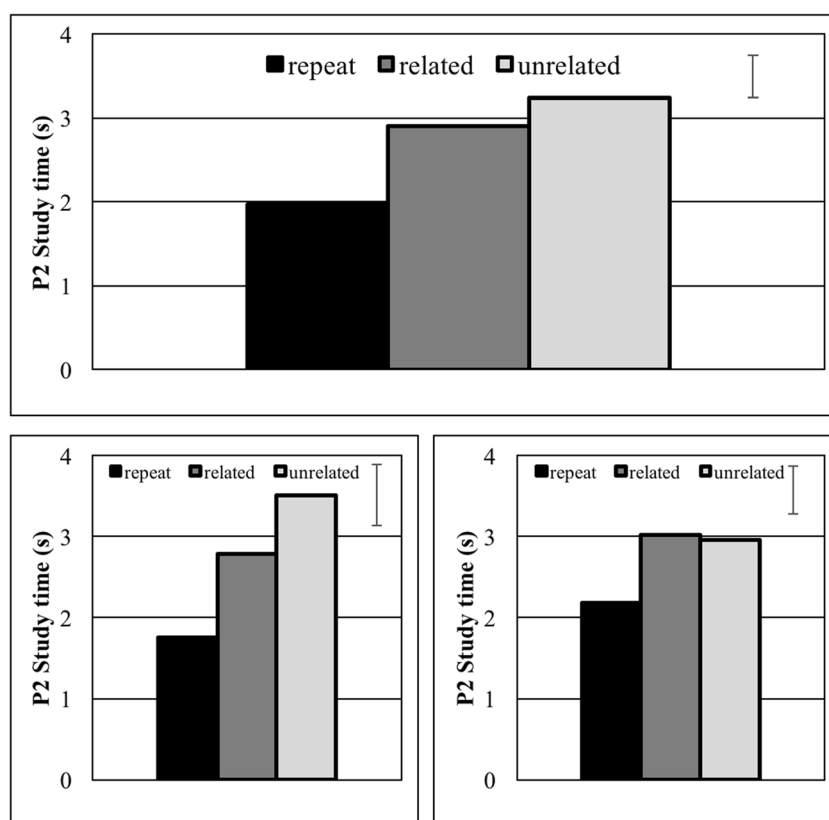


Fig. 1 Experiment 1: P2 study time as a function of condition collapsed across lag (top panel), at a lag of zero (bottom left panel), and at a lag of three (bottom right panel). Error bars shown in the upper-right-hand

corner of each plot show the 95% confidence intervals based on the Subject \times Condition interaction (Loftus & Masson, 1994)

As mentioned above, P1 recognition could not be compared across related and unrelated conditions because P1 item characteristics were confounded with condition. However, there was no confound across lag. If P1 is retrieved during a related P2, but only under massed conditions, then it would follow that related P1s would be recognized more often when massed than when spaced.³ In contrast, lag should have no effect on P1 recognition when the word is unrelated to P2. Indeed, related P1 was recognized more when massed ($M = 0.77$, $SD = 0.19$) than when spaced ($M = 0.71$, $SD = 0.21$), $t(63) = 3.27$, $p = .002$, $d = 0.41$, $r = .71$. Unrelated P1s were recognized at the same rate in the massed condition ($M = 0.67$, $SD = 0.22$) as in the spaced condition ($M = 0.67$, $SD = 0.22$), $t(63) = -0.21$, $p = .834$, $d = -0.03$, $r = .77$.

Prior to introducing the next experiment, we more precisely evaluate the nature of the study-time difference between conditions. Central tendencies of response-time distributions mask many interesting aspects of how response times vary (Van Zandt, 2000), in part due to the skewed nature of such distributions. To our knowledge, there are no published analyses of self-paced study-time distributions, so we adopted techniques from the study of response times more generally.

Analysis of study-time distributions There is an extensive literature on fitting response-time distributions to data from attention and memory tasks (Luce, 1986; Van Zandt, 2000, 2002). Self-paced study is a task in which learners study a stimulus until they choose to proceed. One rudimentary model for such a process is an accumulator with a single-boundary absorbing state: As they study, learners monitor the progress of some latent variable (like memory strength) and terminate study when strength reaches some predetermined boundary. A distribution that both fits response times well and can be meaningfully interpreted within such a framework is the shifted Wald (SW; also called inverse Gaussian) distribution. It is conceptually similar to the popular drift-diffusion model (Ratcliff, 1978), in that it defines a distribution of time taken for a Wiener process (i.e., Brownian motion) to reach a fixed value (Anders, Alario, & van Maanen, 2016; Folks & Chhikara, 1978; Matzke & Wagenmakers, 2009). Unlike the drift-diffusion model, it posits only a single decision boundary (see Anders et al., 2016, for a thorough illustration of this distribution and a discussion of how its parameters relate to the psychological underpinnings of reaction time).

The underlying process assumes a single value X starting at 0 and accumulating with noise at rate γ until reaching boundary α . A third parameter, θ , describes a delay in the onset of

³ We thank an anonymous reviewer for pointing this out.

accumulation and thus can be thought of as all processes external to the accumulation of evidence (including perceptual and response processes). The probability density function is:

$$f(X|\gamma, \alpha, \theta) = \frac{\alpha}{\sqrt{2\pi(X-\theta)^3}} \cdot \exp \left\{ -\frac{[\alpha-\gamma(X-\theta)]^2}{2(X-\theta)} \right\}$$

Like other distributions used in modeling response times, the SW distribution is unimodal and positively skewed. An increase in γ results in a smaller tail, with more of the probability distributed near the mode. An increase in α results in more variation around the mode. An increase in θ corresponds to an overall (positive) shift in the distribution.

For each participant, a distribution was fit to the P2 study time from the repeated, related, and unrelated conditions. In order to maximize the number of observations, the data were first collapsed across lag. The study times were then fit using the data from each lag separately. The data were fit using R (see Anders et al., 2016). Parameter values are shown in Fig. 2. The model fit diagnostics are included in the Appendix; the outcomes of those assessments are in line with the recommendations of Anders et al. (2016).

To evaluate the role of each of the parameters in producing the reduction in mean study time, we developed a novel non-parametric procedure. Because fits to individual subjects can vary widely, the distribution of a given parameter across subjects can be massively skewed. Therefore, it is important to use a technique that does not allow outliers or high-leverage points to have undue influence. We assigned a value of -1 , 0 , or 1 to each parameter-participant pair, corresponding to how well the relative ranking of how closely that parameter's value across conditions matched the shorter mean study time that was observed. For example, it was observed that relatedness reduced study time. Therefore, if γ is driving the reduction in study time, then γ should be the largest for repeated words, followed by related words, and then unrelated words. Such a result would indicate that the relatedness of P1 specifically increases the accumulation rate. In contrast, if α is driving the effect, then the parameter values should be the smallest for repeated words, followed by related words, and then unrelated words. This result would indicate that learners require less information before terminating study, which would be inversely related to the relatedness between P2 and P1. A similar pattern should hold for θ if the reduction in study time is being driven by θ , and would suggest that the relatedness of P1 would reduce the delay before the onset of accumulation. For each parameter, if the ranking is consistent with the reduction in mean study time, a 1 is assigned to that parameter-participant pair. A value of -1 is given if the observed ranking is in the opposite direction, and all intermediate deviations are given a value of 0 . This approach is identical to computing Spearman's rank correlation coefficient for each participant, in

which the covariance is divided by 2 degrees of freedom. Values greater than zero indicate that the rankings lie in the predicted order more often than what would be expected by chance alone.

Rate of accumulation of evidence A one-sample permutation test revealed that the recoded measure of rank was greater than zero when the data were collapsed across lag ($0.35, p = .0001$), at lag of zero ($0.25, p = .0033$), and also at a lag of three ($0.41, p < .0001$). As shown in Fig. 2, γ from the repeated condition was larger than γ from the related and unrelated conditions. This result suggests that evidence is accumulated faster for items that are repeated.

Decision boundary A one-sample permutation test revealed that the recoded measure of rank was greater than zero when the data were collapsed across lag ($0.30, p = .0006$) and at a lag of zero ($0.43, p < .0001$), but not at a lag of three ($-0.11, p = .900$). Here, we can see that the decision boundary was more liberal for related than unrelated words, and yet more liberal for repeated words. Study was terminated with the least evidence for repeated words and with the most evidence for unrelated words. The graded nature of this finding parallels the effects seen on mean study time and supports the argument that related words invite similarly curtailed processing as repeated words (Shaughnessy et al., 1972), but to a less dramatic extent. Like the effect on mean study time, this effect was only apparent during massed presentation.

Extrajudicial processing An analysis of θ revealed no evidence for differences among conditions, except for a weak tendency for spaced presentations of repeated words to elicit less extrajudicial processing ($0.16, p = .039$).

Though these fits are exploratory, they suggest that the reason that a reduction in study time is observed for related P2s and for repeated words is that these items are held to a lower standard for self-paced study than are unrelated P2s. This effect is revealed by differences in actual study time, as well as differences in the inferred decision threshold from an analysis of the distribution of response times. Notably, this result mirrors a now-unpopular theory from the literature on the spacing of repetitions that claimed the basis for the advantage of spaced practice as owing to the diminution of processing under massed conditions (Crowder, 1976; Greeno, 1970; Underwood, 1969, 1970; Waugh, 1970). Interestingly, and unlike repeated words, the rate at which information is collected during study of P2 does not increase when preceded by a related item. These findings align with aspects of our results but provide no means of understanding the advantage in memory for the first member of a pair of related words presented under massed conditions (Tullis et al., 2014). In Experiment 2, we explore the specific relationship between study time and memory, focusing on the origin of the reminding effect.

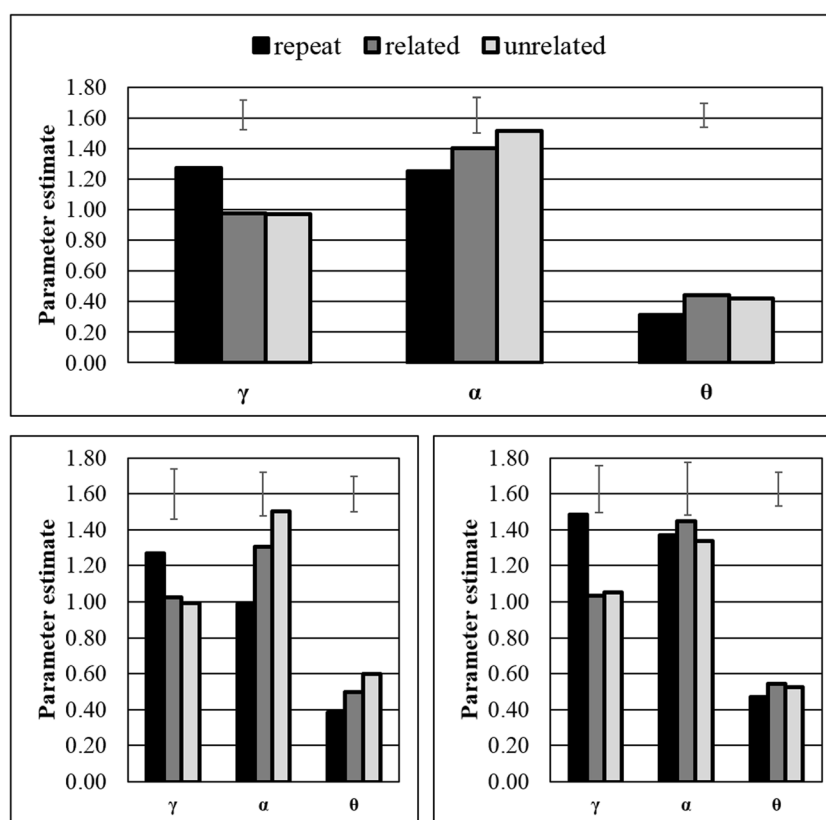


Fig. 2 Experiment 1: Mean of each parameter of the SW distribution for the data collapsed across lag (top panel), at a lag of zero (bottom left panel), and at a lag of three (bottom right panel). Black bars indicate repeated words, dark gray bars indicate related words, and light gray

bars indicate unrelated words. The error bars shown above the bars for each parameter show the 95% confidence intervals based on the Subject \times Condition interaction (Loftus & Masson, 1994)

Discussion

A reduction in P2 study time was observed when that item was a repetition of a prior presentation, replicating previous findings (Shaughnessy et al., 1972; Zimmerman, 1975). The new result here is that a reduction in P2 study time was seen when that word followed a semantic associate. This result suggests that P2 undergoes faster, more efficient, or reduced processing when P1 is related. The fact that this result occurred only following massed presentation of the pair indicates that the potential effect of relatedness is no longer present after a relatively short lag, consistent with what is known about the benefits of relatedness on memory (Tullis et al., 2014). Consistent with prior research, related P2s were recognized as often as unrelated P2s. However, related P1s were recognized more often when massed than when spaced—a pattern that was not observed with unrelated P1s. This pattern is consistent with the study-time data, though criterion effects are also possible.

Although these results are consistent with a view of reminding that treats repetition and semantic association as two points on a continuum of relatedness, the fact that the result obtained only under massed presentation does raise concerns that the reported effects instead reflect sampling error

and our design's opportunities for multiple comparisons. The second experiment provides an opportunity to replicate the central study-time effect, and further to evaluate the degree to which it is related to the downstream memory effect known to occur when testing with cued recall.

Experiment 2

Method

Subjects Seventy-nine introductory-level psychology students for the University of Illinois at Urbana-Champaign participated in exchange for partial course credit. The data from one participant were not used, as that participant did not complete the experiment within the time allotted for the study. Our sample size for Experiment 2 was based on the effect size using P2 study-time data that we observed in Experiment 1. However, because the number of observations per condition was reduced from 16 to six, we estimated a new standardized effect size by taking the effect size from Experiment 1 and resampling six scores from the within-condition observations. This process reduced the effect size from $d = 0.38$ to $d = 0.32$.

Using this estimate, we estimated a need for 79 subjects to achieve statistical power of 0.8.

Design The experiment used a 2 (semantic condition) \times 2 (lag) within-subjects design. Items within a pair were either related or unrelated to each other. Repeated words were not included in order to increase the power of the design to replicate the critical comparison between related and unrelated pairs. Pairs were separated by a lag of either zero or two intervening items.

Materials Thirty-six primary associate pairs were collected from the University of South Florida Free Association Norms database (Nelson et al., 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 extralist 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 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 ($SD = .161$), respectively. All of the unrelated pairs had a forward and backward associative strength of zero. The average test 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., two positions in the study per slot) and two 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. For each participant, each pair was randomly assigned to serve in 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 (e.g., king-queen), and the left-hand member of that pair served as P1 (e.g., king-pepper). Therefore, P2 was controlled across conditions. This procedure yielded 48 words plus two 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 number of times within each of the four study conditions (when collapsed across study list halves). The fillers were not tested.

Procedure The study procedure was identical to Experiment 1 in terms of instructions and timing. After all of the items were presented, participants were given an independent-probe cued-recall test, using extralist cues. The extralist cues were composed of words that were semantically related to the target, the first letter of the target, and a number of dashes equal to the number of remaining letters. 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

P2 study time The critical finding to be replicated from Experiment 1 is the effect of relatedness on study time for P2. Collapsed across lag, P2 study time was less for related pairs ($M = 4.73$ seconds) than for unrelated pairs ($M = 5.70$), $t(77) = -3.27, p = .002, d = 0.41, r = .93$. At a lag of zero, P2 study time was less for related pairs ($M = 4.03$ seconds) than for unrelated pairs ($M = 5.57$), $t(77) = -4.02, p < .001, d = 0.50, r = .88$. At a lag of two, P2 study time was less for related pairs ($M = 5.44$) than for unrelated pairs ($M = 5.83$), but not significantly so, $t(77) = -0.92, p = .362, d = 0.11, r = .88$. The data (collapsed across lag) are displayed in Fig. 3, and completely replicate the results of Experiment 1.

Cued recall performance Collapsed across lag, the reminding effect was present: Cued recall performance of P1 was higher for related pairs ($M = 0.62$) than for unrelated pairs ($M = 0.54$), $t(77) = 3.21, p = .002, d = 0.40, r = .59$. At a lag of zero, cued recall performance of P1 was higher for related pairs ($M = 0.63$) than for unrelated pairs ($M = 0.54$), $t(77) = 3.04, p = .003, d = 0.38, r = .42$. At a lag of two, cued recall performance of P1 was higher for related pairs ($M = 0.60$) than for unrelated pairs ($M = 0.55$), but not significantly so, $t(77) = 1.67, p = .099, d = 0.21, r = .42$.

We also examined memory for P2 even though it is not directly relevant to the reminding effect. Collapsed across lag, cued recall performance of P2 was higher for related pairs ($M = 0.55$) than for unrelated pairs ($M = 0.49$), $t(77) = 3.10$, $p = .003$, $d = 0.38$, $r = .54$. 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 effect was not significant, $t(77) = 1.76$, $p = .082$, $d = 0.22$, $r = .41$. At a lag of two, cued recall performance of P2 was higher for related pairs ($M = 0.54$) than for unrelated pairs ($M = 0.46$), $t(77) = 2.56$, $p = .012$, $d = 0.32$, $r = .33$. The data (collapsed across lag) are displayed in Fig. 4. We do not compare P1 and P2 because of the inherent list position confound between those items.

Study time and cued recall 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, six participants' data were dropped due to perfect separation. That is, for some of the data sets, there existed a study time at 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. If a participant's data from at least one of the two conditions could not be fit, then 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). This measure indicates how much study time is required for a given item in order for it to be recalled with 50% probability. The related condition should yield a *lower* effective rate parameter. The median intercept and slope were computed because there were a few participants that had either a positive or negative slope that was almost equal to zero, yielding massive inflation of the rate parameter and highly skewed distributions. These data are shown in Fig. 5 and Table 2.

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 median effective rate for the conditions (see Ernst, 2004). The observed median difference between conditions was compared with the sampling distribution of differences generated by the data and revealed that the probability of obtaining this difference under the null is unlikely ($p_{\text{two-tailed}} = .016$). In addition, the median slope for the unrelated condition did not significantly differ

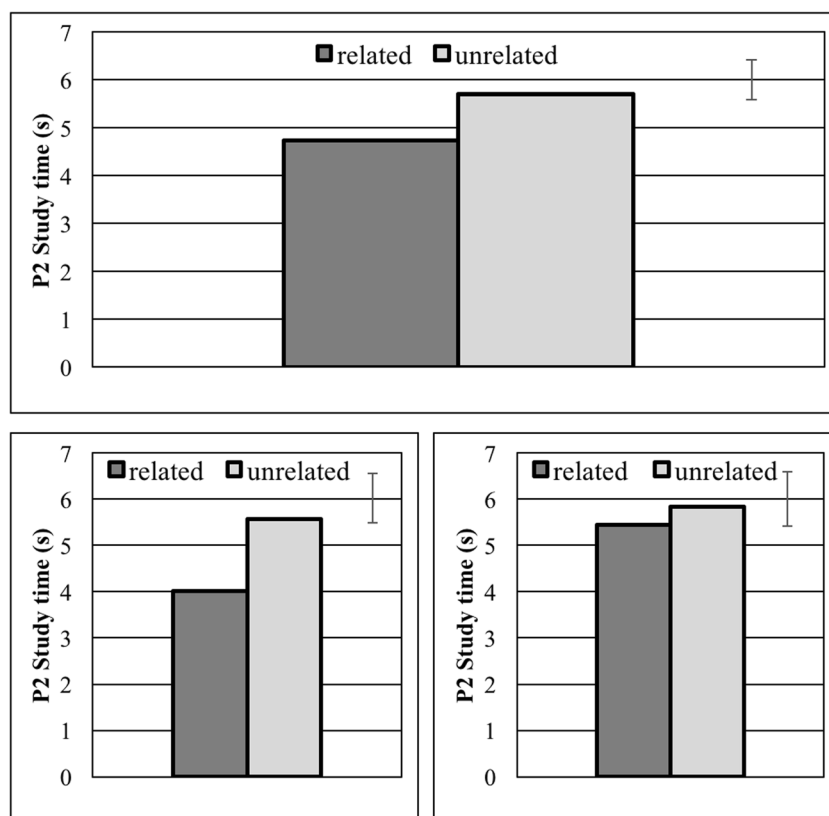


Fig. 3 Experiment 2: Mean P2 study time as a function of condition collapsed across lag (top panel), at a lag of zero (bottom left panel), and at a lag of two (bottom right panel). Dark gray bars indicate related words,

and light gray bars indicate unrelated words. The error bars shown in the upper-right-hand corner of each plot show the 95% confidence intervals based on the Subject \times Condition interaction (Loftus & Masson, 1994)

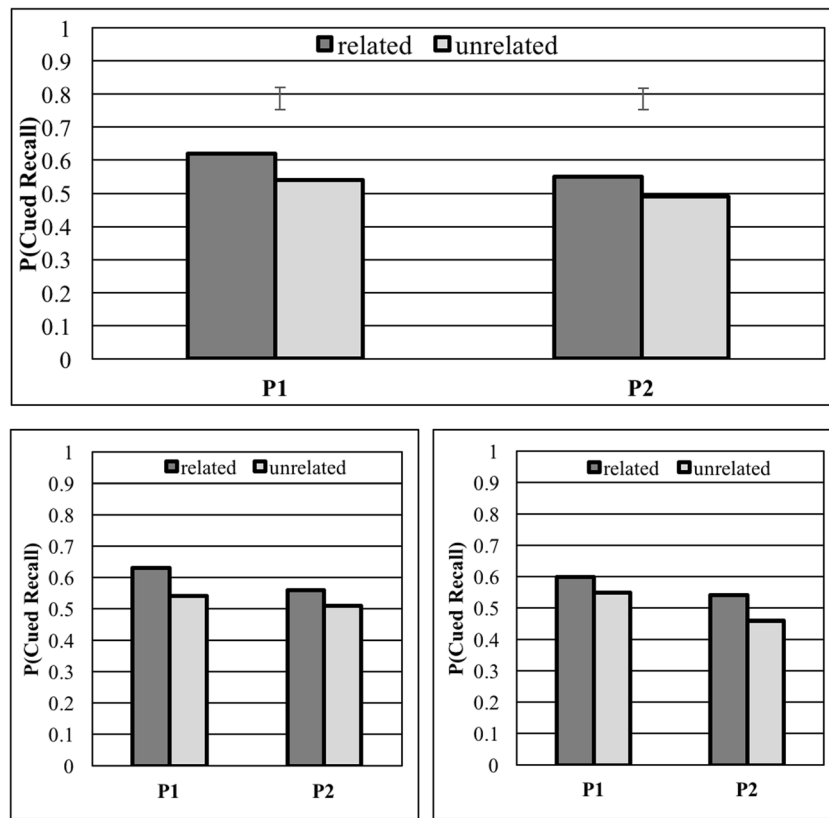


Fig. 4 Experiment 2: Mean cued recall performance for P1 and P2, collapsed across lag (top panel), at a lag of zero (bottom left panel), and at a lag of two (bottom right panel). Black bars indicate related pairs, and

gray bars indicate unrelated pairs. The error bars shown in the above the bars for each item show the 95% confidence intervals based on the Subject \times Condition interaction (Loftus & Masson, 1994)

from zero ($p_{\text{two-tailed}} = .453$), but the median slope from the related condition did ($p_{\text{two-tailed}} = .026$). Taken together, these results indicate that when P2 is related to P1, more P2 study time enhances memory for P1. However, it is important to note that because P1 was not held constant across conditions, this result could reflect item or trial effects. For example, it is possible that when P1 was encoded more thoroughly, P2 study

time was reduced. This hypothesis was evaluated by repeating the above analysis using P1 study time as a predictor of P2 recall, as summarized in Table 3. Here, there was no significant difference between related and unrelated pairs ($p_{\text{two-tailed}} = .340$). This pair of results suggests that the effect of P2 study time on P1 recall was 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., the quality of encoding of P1).

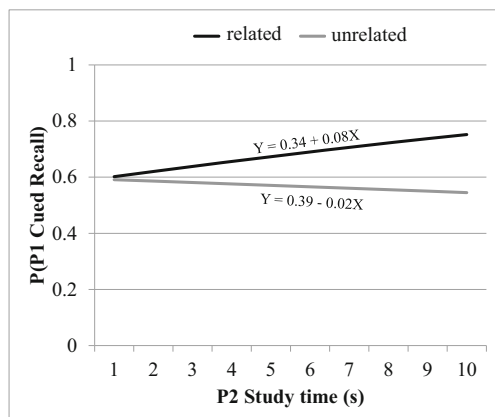


Fig. 5 Experiment 2: Cued recall of P1 as a function of P2 study time, collapsed across lag. The black line represents the related condition and the gray line indicates the unrelated condition. At this scale, the functions appear linear, but they are logistic when plotted over a larger range of study times

Discussion

To summarize, there are three results in this experiment that are consistent with the idea that study reminders influence

Table 2 Experiment 2: Median y-intercepts, slopes, and 50% effective rates for each condition from the logistic fits

	Related	Unrelated	Related	Unrelated
y-intercept	0.34	0.39	0.15	-0.02
Slope	0.08	-0.02	-0.01	0.00
50% effective rate	1.30	2.98	2.56	1.99

Note. The first and second columns show parameter values from the fit of logistic functions predicting P1 recall from P2 study time. The third and fourth columns show parameter values from the fit of logistic functions predicting P2 recall from P1 study time

later memory. First, less study time was allocated to P2 when it followed a semantic associate, replicating Experiment 1. This result is consistent with the notion that retrieval of P1 constrained or facilitated the processing of P2. Second, cued recall performance was higher for P1 when the two items were related, compared with when they were unrelated (i.e., the reminding effect of Tullis et al., 2014). Finally, when the two items were related, memory for P1 was predicted by P2 study time. All of these effects were driven in large part by pairs that were massed.

General discussion

Research on reminding has typically relied on downstream test performance as a basis for inferring the action of reminding. 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 study time measures at the time in which the initial reminding event is thought to occur.

We have shown here that related P2s are studied for less time than unrelated P2s when the lag between P1 and P2 is short. This finding suggests that P1 affected processing of related P2s, but that this effect is of a relatively short duration. We also replicated the finding that repetitions reveal an even more dramatic reduction in study time, also especially at short lags (Shaughnessy et al., 1972), indicating that a common process—here proposed to be *reminding*—underlies study time and memory effects in both cases. An analysis of the shapes of the response-time distributions indicates that the common process here might be a reduction in the strength threshold for cessation of study, though it must also be noted that repetition additionally seemed to increase the rate with which information was amassed in service of study. In Experiment 2, in which we used a cued-recall test, memory performance for P1 was higher when it had been followed by a related P2, replicating the reminding effect (Tullis et al., 2014; Tullis et al., 2014). This boost in memory is thought to result from the memory-potentiating retrieval of P1 in response to a related P2 (Benjamin & Tullis, 2010). The relationship between study time and memory is supported by the finding that study times of related P2 predicted memory for P1 (but that this relation was not evident for unrelated pairs). The fact that this was not observed for unrelated pairs suggests that P1 had a forward effect on P2 study time. Coupled with the finding that related P2s are studied for less time than unrelated P2s, our interpretation is that when P1 and P2 are related, learners compromise P2 study time in the face of uncontrolled retrieval of P1. The net effect of this is to reduce P2 study time and enhance P1 recall. This result also undermines the

notion that the observed reduction in related P2 study time was simply a priming effect, as it is not clear why priming would also enhance P1 memory.

It should be noted that this result is correlational, and although we attempted to rule out one possible explanation, it is impossible to rule out all possible alternative explanations. For example, it is possible that when P1 is processed more thoroughly, P2 study time is reduced. In that explanation, the enhancement in memory to P1 is a cause, not an effect. We cannot definitively rule out this possibility. However, we can say that P1 study time was *positively* related to P2 study time (average correlation = 0.34), and, so, we can rule out the possibility that when P1 is processed more thoroughly, P2 study time is reduced.

Although none of these reminding effects were observed at significant levels at the longer lags in our experiments, there were a few characteristics about the current design that made reminding less likely overall. Most importantly, participants were given complete control over the duration of their study. It is possible that participants did not allocate enough study time to P1 in order to support reminding at longer lags. Second, participants were given an independent-probe cued-recall test in order to reduce potential opportunities for reminding at test. As a result, this forced participants to use an experimenter-generated cue for interrogating memory, rather than their own. This design choice may lessen the benefits of idiosyncratic mediators developing during reminding.

Results from the current experiments suggest that the consequences of reminding are not limited to test performance, but can also influence more immediate behavior that is observed during the reminding event. Specifically, a reminding event is more likely to elicit retrieval, which influences behavior in response to that event. Such behavior is 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 may subserve metacognitive monitoring (Benjamin & Bjork, 1996; Benjamin, Bjork, & Schwartz, 1998; Tullis et al., 2014, Experiment 3b). Specifically, Tullis et al. (2014) observed inflated JOLs in response to a reminding event. Additionally, P2 JOLs predicted memory for related but not for unrelated P1s. Here, we have presented evidence that extends the role of retrieval to the control of study time (see also Son & Metcalfe, 2000).

A related result is the *testing effect*, which demonstrates that retrieval is a potent tool for promoting retention (Benjamin & Pashler, 2015; Karpicke & Roediger, 2007, 2008; Tullis, Finley, & Benjamin, 2013). Both the reminding effect and the testing effect are results that blur the boundaries between aspects of memory that are sometimes presumed to

be separate. The testing effect demonstrates that retrieval involves additional encoding, and the reminding effect demonstrates that encoding involves retrieval. 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 complex cognitive tasks. Theories that postulate specific “modes” that one must enter for retrieval, for example, are incompatible with this view (cf. Nyberg et al., 1995). It is worth remembering this fact when considering, for example, the function of teaching—which should not simply be about transmitting new information but 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 retrieval of information that may interfere with future learning (Ross, 1984, 1987).

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 one’s memory and knowledge structures. Similarly, when an event reminds us of an earlier event, this can influence how we behave in response to that current event, permitting flexibility in one’s behavior and decision-making.

Author note Geoffrey L. McKinley, Aaron S. Benjamin, and Brian H. Ross, Department of Psychology, University of Illinois at Urbana-Champaign.

A portion of this research was conducted for the first author’s master’s thesis at the University of Illinois at Urbana-Champaign.

Appendix

Table 3 Mean P1 study time for each Lag × Condition combination, for Experiment 1 and 2

Experiment 1	Experiment 2	
	Massed	Spaced
Related	3.39 (3.20)	3.31 (2.97)
Unrelated	3.99 (3.90)	3.57 (3.11)
Repeated	3.70 (3.64)	3.34 (3.30)

Standard deviations are in parentheses

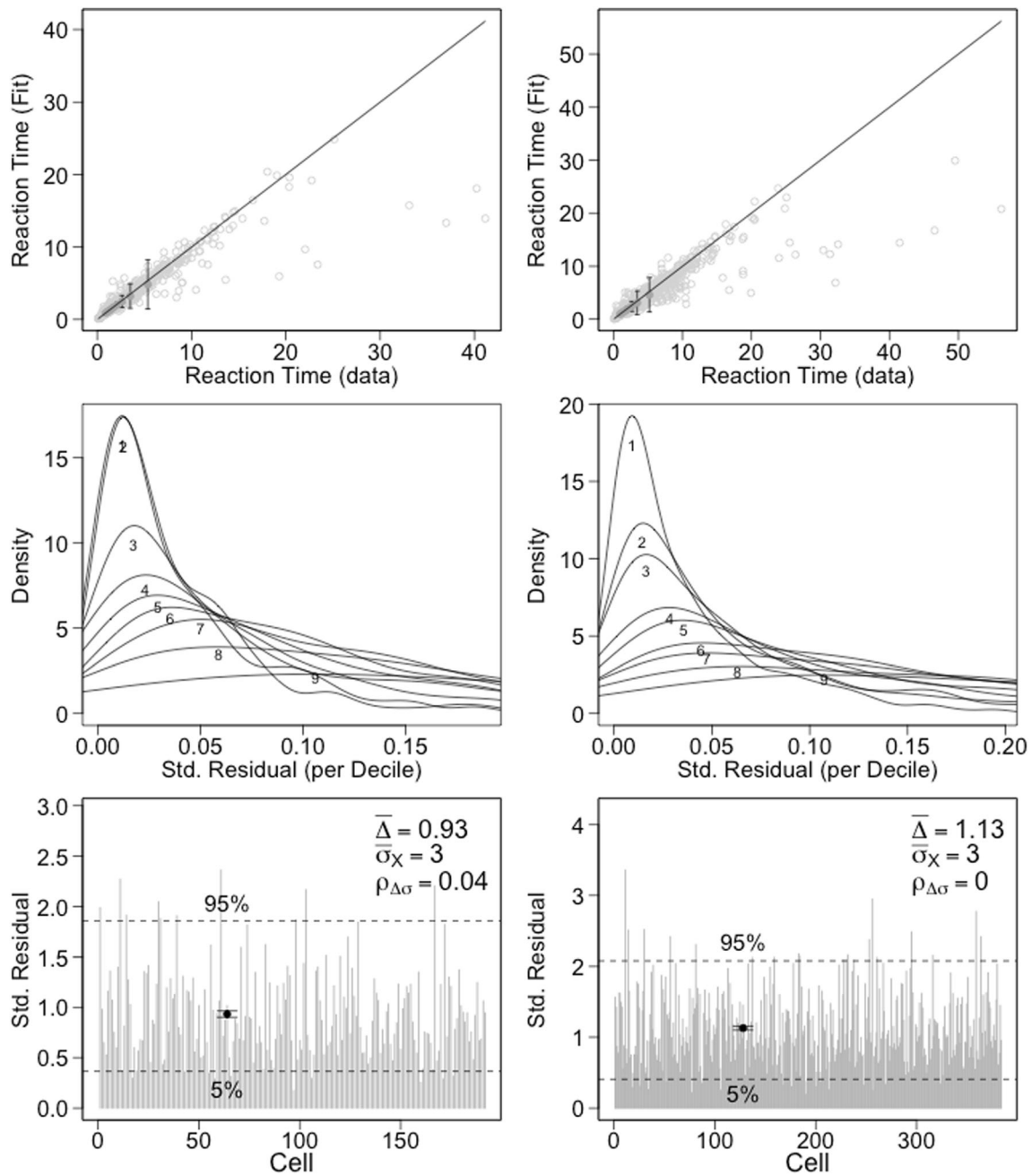


Fig. 6 Model fit diagnostics for fitting the shifted Wald distribution to P2 study time. The top row contains quantile-quantile (QQ) plots, in which the observed study time deciles are compared against the model-predicted deciles (in seconds) for each of the unique experimental design cells. The middle row contains the by-decile distribution of the standardized residuals. These were computed by standardizing the difference between the

observed data deciles and the predicted deciles for each cell. The bottom row contains the summed standardized residuals for each cell, across the nine deciles. The left column provides these diagnostics for the data collapsed across lag, or each Subject × Condition cell. The right column provides the diagnostics for each Subject × Condition × Lag cell

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Anders, R., Alario, F., & van Maanen, L. (2016). The shifted Wald distribution for response time data analysis. *Psychological Methods, 21*(3), 309–327.
- 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., & 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., & 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.
- Bruce, D., & Weaver, G. E. (1973). Retroactive facilitation in short-term retention of minimally learned paired associates. *Journal of Experimental Psychology, 100*, 9–17.
- Challis, B. H. (1993). Spacing effects on cued-memory tests depend on level of processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(2), 389–396.
- Cramer, J. S. (1991). *The logit model: An introduction for economists*. London, UK: Edward Arnold.
- Crowder, R. G. (1976). *Principles of learning and memory*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Ernst, M. D. (2004). Permutation methods: A basis for exact inference. *Statistical Science, 19*(4), 676–685.
- Folks, J. L., & Chhikara, R. S. (1978). The inverse Gaussian distribution and its statistical application. *Journal of Royal Statistical Society of Great Britain, 40*, 263–289.
- Fraundorf, S. H., Watson, D. G., & Benjamin, A. S. (2015). Reduction in prosodic prominence predicts speakers' recall: implications for theories of prosody. *Language, Cognition and Neuroscience, 30*(5), 606–619.
- Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology, 12*, 306–355.
- 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.
- Glanzer, M. (1969). Distance between related words in free recall: Trace of the STS. *Journal of Verbal Learning and Verbal Behavior, 8*, 105–111.
- Greene, R. L. (1989). Spacing effects in memory: Evidence for a two-process account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 371–377.
- Greeno, J. G. (1970). Conservation of information-processing capacity in paired-associate memorizing. *Journal of Verbal Learning and Verbal Behavior, 9*, 581–586.
- 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, UK: 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.
- Jacoby, L. L. (1974). The role of mental contiguity in memory: Registration and retrieval effects. *Journal of Memory & Language, 13*(5), 483–496.
- Jacoby, L. L., & Wahlheim, C. N. (2013). On the importance of looking back: The role of recursive reminders in recency judgments and cued recall. *Memory & Cognition, 41*, 625–637.
- 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.
- Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review, 1*, 476–490.
- Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. New York, NY: Oxford University Press.
- Matzke, D., & Wagenmakers, E.-J. (2009). Psychological interpretation of the ex-Gaussian and shifted Wald parameters: A diffusion model analysis. *Psychonomic Bulletin & Review, 16*, 798–817.
- 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.
- Nyberg, L., Tulving, E., Habib, R., Nilsson, L.-G., Kapur, S., Houle, S., . . . McIntosh, A. R. (1995). Functional brain maps of retrieval mode and recovery of episode information. *Neuroreport, 6*, 249–252.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*, 59–108.
- Robbins, D., & Bray, J. F. (1974). Repetition effects and retroactive facilitation: Immediate and delayed test performance. *Bulletin of the Psychonomic Society, 3*, 347–349.
- Ross, B. H. (1984). Reminders 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 reminders. *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.

- Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *26*, 204–221.
- Tullis, M., 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). Reminders influence the interpretation of ambiguous stimuli. *Psychonomic Bulletin & Review*, *21*, 107–113.
- 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.
- Van Zandt, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin & Review*, *7*(3), 424–465.
- Van Zandt, T. (2002). Analysis of response time distributions. In J. T. Wixted (Ed.), *Stevens' handbook of experimental psychology* (3rd ed., pp. 461–516). San Diego, CA: Academic Press.
- Wahlheim, C. N., & Jacoby, L. L. (2013). Remembering change: The critical role of recursive reminders in proactive effects of memory. *Memory & Cognition*, *41*, 1–15.
- 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.
- Zimmerman, J. (1975). Free recall after self-paced study: A test of the attention explanation of the spacing effect. *The American Journal of Psychology*, *88*, 277–291.