Item repetition and retrieval processes in cued recall: Analysis of recall-latency distributions

Yoonhee Jang¹ · Heungchul Lee²

Published online: 8 February 2019
© The Psychonomic Society, Inc. 2019

Abstract

The SAM (search of associative memory) model provides a unified account of accuracy effects, assuming that retrieval is a cue-dependent two-stage process of sampling and recovery, which depends on the strength of items relative to all others and on that item associated with the sampling trace, respectively. On the other hand, the relative strength model uniquely provides latency predictions, assuming that recall latency is determined solely by relative strength (similar to the sampling rule in SAM): Latency should remain unchanged for strong and weak items in pure lists, but will be shorter for strong items than for weak items in mixed lists. To test the predictions, the present study examined accuracy and latency distributions, which were fit with the ex-Gaussian, using item repetition as a means of strengthening. Massed versus spaced repetitions were used where repetitions were either cue–target pairs or cue alone. When repetitions were spaced in mixed lists, accuracy and latency both increased with cue–target repetitions, relative to cue-only repetitions, and slow recall for cue–target repetitions was due to initially nonretrievable items. However, even after successful recall on a pretest, cue–target repetitions led to an increase in latency in pure lists. These findings are difficult to reconcile with relative-strength explanations of latency. They indeed suggest that (1) separate traces are created for each repetition, (2) memory traces are updated if the item is retrieved (otherwise, new traces are stored), and (3) recovery plays a role in latency, which are discussed with the distinction between sampling and recovery of SAM.

Keywords Item repetition · Recall latency · Sampling and recovery

Atkinson and Shiffrin (1968) produced a model framework that is often referred to as the modal model of memory, based on a distinction between structural features of the memory system and control processes. Structural features include the separate stores of sensory registers, short-term store (STS), and long-term store (LTS). Control processes include the operations to control memory, such as coding, rehearsal, and retrieval strategies. The framework made it possible not only to analyze memory search and retrieval processes but also to develop detailed models for specific experimental tasks, such as the search of associative memory (SAM) model, initially developed for free recall (Raaijmakers & Shiffrin, 1980, 1981) and later generalized to cued recall and recognition (Gillund & Shiffrin, 1984). SAM has been restricted for the most part to accuracy

predictions, but it can be a useful tool in illuminating all forms of relationships between accuracy and latency. To examine this possibility, in the present study, we investigated how item repetition affects accuracy and latency in a cued-recall paradigm. We particularly chose item repetition because even if recall accuracy mostly increases with item repetition, recall latency can decrease or increase (e.g., Bousfield, Sedgewick, & Cohen, 1954; Unsworth, 2015; Wixted, Ghadisha, & Vera, 1997). Additionally, we analyzed recall-latency distributions because, despite their usefulness, such analyses of cued-recall latency are rare (except for Nobel & Shiffrin, 2001, and its companion article by Diller, Nobel, & Shiffrin, 2001), compared with those of free-recall and recognition latencies.

The recall of items from long-term memory is often viewed as a competitive process in which an item's chance depends not only on its own strength but also on the strengths of other items (e.g., Shiffrin, 1970). Based on the search model of Shiffrin, SAM posits that memory traces are represented in LTS as separate events, which contain three types of information: item, associative, and context information. Item information refers to features of the item content. Associative information represents relationships between traces in memory. Context

✉ Yoonhee Jang
yoonhee.jang@umontana.edu

¹ Department of Psychology, University of Montana, 32 Campus Drive, Missoula, MT 59812-1584, USA

² Net Intelligence & Research, Seoul, South Korea

information refers to a special case of associative information in which the associate is the context. Retrieval cues (e.g., a test item with context) placed in STS serve as the probe of long-term memory and activate a search set consisting of the entire set of traces. Recall proceeds as a sequential search with cycles of sampling and recovery. In each cycle of the search, a trace is sampled with replacement, and the probability of sampling the trace depends on its relative strength, compared with the summed activation of all traces in LTS. Then, an attempt is made to recover as much information as possible from the sampled trace. If the desired trace is sampled, and then if enough information is recovered, a correct response occurs. Even after an item's trace is sampled, the item cannot be recalled if it is not successfully recovered. Recovery of an item depends on its absolute strength, which does not affect sampling. In other words, SAM assumes that retrieval from LTS is a cue-dependent two-stage process, including the sampling of a trace based on match to context followed by an attempt to recover the item associated with the sampled trace.

The SAM model has been well able to account for long-term memory effects in terms of accuracy, rather than latency. Nonetheless, the distinction between sampling and recovery can help to explain differences in recall latency as well as accuracy across manipulation conditions for specific experimental tasks. In addition, understanding under which situation dissociations between recall accuracy and latency occur will help memory models, including SAM, provide a better understanding of memory processes and facilitate future progress. Although the two measures of recall accuracy and latency appear to be separate, they are related to each other and to some memory models as well, and we consider these models (including SAM) in various situations.

Strengthening manipulations and sampling

In fact, Rohrer (1996) and Wixted et al. (1997) proposed a relative strength model, examining recall latency in a free-recall paradigm, which is the kind of model that forms the core of SAM's account, including the sampling (relative strength) and recovery (absolute strength) process. Wixted et al. (1997) referred to SAM as the best-known example of relative strength models, and their model often provided interpretations similar to those of the SAM model. However, the relative strength model is essentially different from SAM, uniquely predicting performance on recall latency. The key assumption of the relative strength model is that recall latency is determined solely by relative strength (which is similar to the sampling rule in SAM). There has been some evidence for (and against) the relative strength model that various experimental manipulations do or do not affect recall latency.

For example, list length and strength are frequently used to evaluate differences in performance between conditions. List

length is inherently manipulated in a pure-strength situation, such as a long list versus a short list. Rohrer and Wixted (1994) showed that as list length increased, recall accuracy decreased (although the number of items recalled increased) and recall latency increased. According to the SAM model, in a long list, more items should be matched to the cue, and the relative activation associated with the trace decreases, which results in greater interference because each new item introduces more variability. According to the relative strength model, relative strength decreases with an increase in the size of the search set as other items are added. That is, both models suggest that when list length is manipulated, changes in both recall accuracy and latency are due to the relative differences between sampling strengths (despite somewhat different interpretations in detail).

On the other hand, there are various ways to strengthen items, and a strength manipulation can be achieved in a pure-strength or mixed-strength (e.g., a list including both strong and weak items) situation. According to the relative strength model, the relationship between recall accuracy and latency depends on the type and method of strengthening. For example, Rohrer and Wixted (1994) found that in pure lists, strengthening words on a list by allowing extra study time increased recall accuracy but did not affect recall latency. In the SAM model, longer study times increase the strength between studied items and memory traces, enhancing sampling strengths of the studied items, and as the retrieval probability for each item is determined by the absolute magnitude of the sampling strength, recall accuracy increases. In the relative strength model, longer study times do not result in a greater number of items within the search set but increase the likelihood that an item within the search set will be recoverable. That is, according to the relative strength model, when presentation duration is manipulated, there is no change in latency because sampling strength remains constant, whereas there is a change in recall accuracy due to the relative differences between recovery strengths. In addition, item repetition (or specifically, the number of presentations) is a common method of strengthening. The relative strength model predicts that in a mixed list, strong items will be recalled faster than weak items because relative strength is greater for strong items than for weak items. Wixted et al. (1997) found evidence for the relative strength model when presenting items once or twice, even if repetitions were sometimes massed and sometimes spaced in the mixed-strength situation. This indicates that when items are strengthened by repetition in mixed lists, a change in latency is due to the relative differences between sampling strengths, which is consistent with the account of accuracy results in the SAM model (strong items are sampled preferentially).

Interestingly, in pure lists, the prediction of the relative strength model is that item repetition should not influence recall latency, similar to presentation duration, because when the memory traces are strengthened by item repetition, all of the competitors are strengthened as well, and an item's relative strength does not change with item repetition even though its

absolute strength does. However, Wixted et al. (1997) found a small but significant increase in recall latency as the number of presentations increased. Also, Bousfield et al. (1954) reported repeated items yielded a large increase in recall latency. Recently, using massed and spaced repetitions, Unsworth (2015) showed that in the massed condition, item repetition did not change recall latency, indicating there was a mixture of massed and spaced repetitions in Wixted et al. By contrast, in the spaced condition, Unsworth found that item repetition increased recall latency, replicating the finding of Bousfield et al. To account for these findings, Unsworth suggested that each presentation with spaced repetitions creates a new copy, and because multiple copies of the same item increase the size of the search set, recall is slower.

It has long been a question whether each repetition of an item yields a new, separate trace or serves to update the old memory trace. The original version of SAM is an example of the models that posit separate storage. Gillund and Shiffrin (1984) wrote, “It seems more likely that n spaced presentations would give rise to n separate images, each with a (somewhat) independent variability contribution and each with differing interitem strengths to different items” (p. 48), suggesting that massed repetitions would be similar to increasing presentation duration, but that spaced repetitions would yield separate traces. However, subsequent work (e.g., Ratcliff, Clark, & Shiffrin, 1990) failed to find support for the separate storage hypothesis, and Shiffrin, Ratcliff, and Clark (1990) provided a variant of the SAM model, assuming that repetitions are accumulated into a single memory representation of increasing strength. The accumulated storage assumption has been successfully applied to later models as well, such as REM (retrieving effectively from memory; Shiffrin & Steyvers, 1997). In this sense, slower recall for spaced-repetition items found in prior studies can be thought of as a challenge to accumulated storage models, including SAM and REM.

It is of great importance to note that regardless of whether a second presentation of an item strengthens the original representation or creates another copy of it, the prediction of the relative strength model is the same. That is, recall latency should remain unchanged, because the probability that a given item will be sampled in the search set does not change (see, Wixted et al., 1997, p. 525, for details). In either case, the relative strength model faces problems with strengthening manipulations by item repetition, and another situation follows.

Retrieval practice and recovery

Even more interesting is the experimental paradigm in which two forms of item repetition are compared. One is restudy or study practice after initial learning during which each item is repeated, similar to spaced repetitions, and the other is retrieval or test practice during which participants receive an intervening

test. An item that is retrieved on the intervening test is found to be strengthened, as seen in performance on a later test, and this phenomenon is known as the testing effect (for a review, see Roediger & Karpicke, 2006). A powerful demonstration compares restudy versus test practice in the absence of feedback, as they affect an immediate final test versus a delayed final test (e.g., 5 min vs. 24 hours later), and in this situation, a crossover interaction is often found. That is, benefits of restudy occur for an immediate final test, whereas benefits of test practice occur for a delayed final test (e.g., Jang, Wixted, Pecher, Zeelenberg, & Huber, 2012; for a review, see Roediger & Karpicke, 2006). A fundamental question in the majority of testing effect studies concerns whether there is a benefit to accuracy from test practice. However, in the present study, we consider theoretical accounts of the relationship between recall accuracy and latency across experimental conditions (e.g., restudy and test practice), using existing models, such as the SAM and relative strength models, rather than those of the long-term accuracy advantage of test practice (for this purpose, see, e.g., Delaney, Verkoeijen, & Spigler, 2010). We focus, therefore, on only one situation in which performance is assessed on an immediate final test.

A few recent studies (e.g., Hopper & Huber, 2018a; van den Broek, Segers, Takashima, & Verhoeven, 2014) reported dissociations between recall accuracy and latency, comparing restudy versus test practice. For example, Hopper and Huber found that in a blocked design, free-recall paradigm (which is equivalent to a pure-strength situation), restudy boosted accuracy on the immediate final test, but test practice decreased latency although it led to lower accuracy. That is, in pure lists, strong items in the restudy condition were recalled more slowly than were weak items in the test practice condition. This result is consistent with that of prior research (Bousfield et al., 1954; Unsworth, 2015)—spaced repetitions in pure lists yielded an increase in recall latency, which is difficult to reconcile with the relative strength model, as it predicts no change in recall latency.

In addition, van den Broek et al. (2014) found that, using cued recall, test practice decreased recall latency on the immediate final test, compared with restudy, whereas recall accuracy was greater with restudy than with test practice. Importantly, first, they did not manipulate restudy versus test practice in a pure-strength situation, as the manipulation took place in a single practice phase and participants were tested on a random selection of items from each condition on the final test. In other words, in a mixed-strength situation, strong items in the restudy condition were recalled more slowly than were weak items in the test practice condition. This pattern is in the opposite direction to the prediction of the relative strength model. Second, unlike Hopper and Huber, van den Broek et al. measured perceived retrieval success during practice, and their participants were not asked to explicitly (or overtly) retrieve items in the test practice condition. Specifically, during practice, items were presented before they were replaced by a prompt to make a retrieval success judgment, and the

participants answered the question “Did you already know the target word?” with “Yes” or “No” after each practice trial. This indicates that retrieval practice is important in its own right to decrease recall latency even if it occurs as covert retrieval of the target. In fact, Jang, Lee, and Huber ([in press](#)) suggested that covert retrieval practice as well as overt retrieval practice (by virtue of test practice) leads to faster recall. Specifically, when participants were asked to indicate the likelihood that the target would be recalled on a subsequent recall test, recall latency on the test was shorter if their judgments were made some time after study (as in the spacing of spaced repetitions). This finding indicates that participants engage in covert retrieval practice when making the prediction, which facilitates access to the item-specific information and therefore results in faster recall.

A possible account of these findings is that a manipulation condition may influence recall latency while having little to do with search set size (in the relative strength model), or it may selectively affect the recovery process for individual items rather than the sampling of a trace based on match to context (in the SAM model). Then, recall latency is not necessarily shorter for strong items than for weak items in mixed lists. MacLeod and Nelson ([1984](#)) suggested that recall accuracy measures how sufficiently an item is stored or encoded for later retrieval, whereas recall latency measures how many steps or stages are necessary to decode the item during retrieval. If retrieval practice reduces decoding steps, then recall latency will decrease regardless of the level of accuracy. In this regard, retrieval practice may not affect search set size or sampling strength as much as study practice. Then, recall latency will be shorter in the test practice condition than in the restudy condition, no matter whether the conditions are presented in pure or mixed lists, and even when recall accuracy is lower in the test practice condition than in the restudy condition. It is possible that recall latency depends not only on the relative differences in sampling strengths (as in the key assumption of the relative strength model) but also on the relative differences in recovery strengths. This possibility is by no means contradictory to SAM, although it is not derived directly from the model.

Overview of the experiments and measuring recall latency

The primary goal of the current study was not only to examine the relationship between recall accuracy and latency across manipulation conditions but also to understand search and retrieval processes during recall. Three experiments were designed, using item repetition as a means of strengthening in a cued-recall paradigm, to test whether strong items are recalled faster than weak items in a mixed-strength situation (Experiments 1 and 2), and whether recall latency remains unchanged in a pure-strength situation (Experiment 3).

Importantly, the aim of these experiments was to explore a parsimonious theoretical account of retrieval, using a more complex model, such as SAM, that allows for variable item strengths, interitem associations, and context strengths.

The question of interest is whether recall latency is affected when item repetition produces superior memory in one condition over another. In the present study, regardless of how item repetition was manipulated, the number of unique items for each condition (including the control condition) was held constant, and each experimental condition consisted of the same number of repetitions to avoid confounding length and strength effects. For the first two experiments, massed versus spaced repetitions were manipulated in which repetitions were either cue–target pairs or cue alone. Compared with the manipulation of massed and spaced repetitions, cue–target versus cue-only repetitions have not been much used. However, this particular comparison was adopted because it allowed us to manipulate the availability of information about the target at the time of the second presentation. Specifically, in the cue–target condition, the same item of the cue and target pair was fully available (or restudy), whereas in the cue-only condition, partial information, or the cue alone was available, and the target would not be unless it was covertly retrieved. In this procedure, the cue–target condition would lead to higher accuracy on average, and whether recall would be faster in the cue–target condition was examined. For the last two experiments, a pretest trial was included before (Experiment 2), or after (Experiment 3) the manipulation took place. In Experiment 2, this procedure allowed us to analyze data for retrievable and nonretrievable items, as prior research suggested that retrievability plays an important role in benefits of restudy (or cue–target repetitions) versus test practice (Jang et al., [2012](#)). This is also important because recall latencies are compared only for those items that were recalled (although error latencies are reported as well). Hence, Experiment 2 additionally examined whether retrievability would affect latency. In Experiment 3, where the cue–target condition was compared with the control condition, the procedure allowed us to examine whether strong items (after restudy) would be recalled slowly or not, even after these items were retrieved on the pretest.

To specify retrieval processes in long-term memory, recall-latency distributions were fit with ex-Gaussian distributions, which have been widely used for free recall and recognition, but not for cued recall. The ex-Gaussian consists of three parameters: the normal’s mean (μ) and standard deviation (σ), and the mean of the exponential (τ) for two serial, independent stages of a response process. The first two represent the mean and standard deviation of a Gaussian stage of retrieval. This stage reflects a brief initiation during which participants shift attention to recalling words or automatic residual processing (Hockley, [1984](#); Rohrer & Wixted, [1994](#)). The last parameter, the skew of the ex-Gaussian, represents the exponentially distributed stage of retrieval or strategic decision processing (Heathcote, Popiel, & Mewhort, [1991](#); Ratcliff & Murdock,

1976; Rohrer & Wixted, 1994; Rouder, Lu, Speckman, Sun, & Jiang, 2005). The sum of μ and τ necessarily equals the total mean latency. Typically, the initiation time (μ) is not critical (i.e., no or little change across conditions), and the term, mean latency or a measure of decision processing, refers to τ (Heathcote et al., 1991; Rohrer & Wixted, 1994). In sum, the ex-Gaussian is fit in order to obtain a more accurate measure of the exponential decline in latency distributions, and the key data reported below are estimates of the third parameter (τ), or the mean of the exponential stage, which represents an increasing number of slow responses forming the right tail of the distribution.

Experiment 1

Experiment 1 aimed to investigate recall latency for strong and weak items using a standard manipulation of repetition type (massed and spaced repetitions) and item type (cue–target and cue-only repetitions) in a single study–test session (i.e., a mixed list). Unsworth (2015) used massed and spaced repetitions in pure lists, showing an increase in recall latency only with spaced repetitions. That is, the main concern in his study was whether recall latency increased with the number of presentation in the massed and spaced conditions, separately. In a mixed list, the relative strength of a strong item is greater than that of a weak item, and if recall latency is determined solely by relative strength, then strong items (such as spaced-repetition items) should be recalled more quickly than weak items (such as massed-repetition items). This prediction has not been examined, and the first experiment reported here was designed to do so. In addition, in the massed condition, information concerning the target of a pair will be mostly available in short-term memory at the time of the second presentation, even when only the cue is presented in the absence of the target, and there will be no benefit of cue–target repetitions over cue-only repetitions. By contrast, in the spaced condition, information concerning the target of a pair will be no longer available for cue-only repetitions unless it is retrieved from long-term memory, and there will be no additional study opportunity, compared with cue–target repetitions. In this manner, recall latencies for strong and weak items are compared between spaced versus massed repetitions particularly in the cue–target condition, and between cue–target versus cue-only repetitions particularly in the spaced condition.

Method

Participants Sixty-six undergraduate students at the University of Montana were recruited and received credit for psychology courses in return for their participation.

Design A 2 (repetition type: massed and spaced) $\times$ 2 (item type: cue–target and cue only) within-subjects factorial design was used.

Materials The stimuli were 60 moderately high-frequency (an average frequency of 60 per million; Kucera & Francis, 1967), singular noun word pairs, from four to seven letters in length. The two words of a pair were semantically and phonologically unrelated. The first 12 pairs constituted practice, and the remaining 48 pairs comprised two blocks of 24 pairs per block and were the only pairs that were analyzed. For each block, six pairs were randomly assigned to each of the four (massed cue–target repetition, massed cue-only repetition, spaced cue–target repetition, and spaced cue-only repetition) conditions.

Procedure Figure 1 shows the experimental procedure of all experiments. Each experiment consisted of study and test phases in a single session, and participants were tested individually. During the study phase, participants were asked to learn word pairs for a later test in which they would be asked to type the second word (target) when given the first word (cue). Each pair was presented as a study item in the center of the screen for 4 s, and 750 ms of blank screen separated the items. All study items were randomly ordered anew for each participant with the restriction that the four conditions were each assigned six pairs in each of two successive blocks of 24 pairs. For an item, either the cue–target pair (in the cue–target condition) or the cue alone (in the cue-only condition) was repeated for 2 s. In the massed condition (e.g., “A-B” in Fig. 1), repetitions occurred in succession, and in the spaced condition (e.g., “C-D”), repetitions occurred after the entire items per block were presented.

During the test phase, participants were instructed to type the target word from each pair when given the cue word. They typed in their responses and pressed the “enter” key after each response to proceed to the next test trial, and, if they had no guess, then they typed “next”. No feedback was given. The test order consisted of the first 24 studied items in randomized order, followed by the next 24 studied items in randomized order. Recall latency for each response was measured as the duration between the presentation of the cue and the first key press of the participant’s response (see also Hopper & Huber, 2018a; Jang et al., *in press*).

Results and discussion

Each response was scored as correct, an intrusion, or no response (i.e., typing “next”). The dependent measures of interest are proportion correct and correct recall latency. The proportions and latencies of intrusions and failures to respond are included for completeness, as reported in Table 1. Results are reported in three sections for each experiment: (1) recall response probability; (2) recall latency in which the latencies for correct and incorrect responses were transformed using the

Fig. 1 Example of the procedure of all experiments. Each experiment consists of study and test phases. Word pairs of “A-B” and “C-D” are presented in the massed and spaced conditions, respectively, of Experiments 1 and 2, and both in the spaced condition of Experiment 3. After the first “A-B,” the second “A-B” represents the cue–target repetition, and the first “A-” represents the cue-only repetition in Experiment 1. The procedure of Experiment 2 is identical to that of Experiment 1,

except that Experiment 2 additionally includes the control condition (no repetition illustrated by the black box) and the pretest (i.e., the first “A-” and “C-”) trial before the manipulation. In Experiment 3, the pretest trial takes place after the manipulation. Note that only for simplicity, the fixed order of presentation for both phases (i.e., “A-B” followed by “C-D”) is illustrated (see text for details)

natural logarithm; and (3) recall-latency distributions fit with ex-Gaussian distributions for correct responses.

Throughout, statistical significance was determined with an alpha of .05 unless otherwise noted, and, for simplicity, we report the main results in the text. The complete results from statistical analyses of Experiment 1 are reported in Appendix A (Tables A1 and A2).

Recall response probability Figure 2 shows mean proportions correct recall (Fig. 2a). A 2×2 analysis of variance (ANOVA) was conducted (Table A1: Correct) with repetition type (massed and spaced) and item type (cue–target and cue only). The main effects were significant, which showed that recall accuracy was higher in the spaced condition than in the massed condition; and recall accuracy was higher for cue–target repetitions than for cue-only repetitions. The interaction was significant, and simple effect analyses (Table A2: Correct) showed that in the massed condition, recall accuracy was not different between cue–target and cue-only repetitions, whereas in the spaced

condition, it was higher for cue–target repetitions than for cue-only repetitions. In addition, for cue–target repetitions, recall accuracy was higher in the spaced condition than in the massed condition. However, the opposite was true for cue-only repetitions: Recall accuracy was marginally higher in the massed condition than in the spaced condition.

We replicated and extended the spacing effect on recall accuracy that an additional spaced learning of the cue–target pair improves the accuracy of long-term memory, whereas there is no additional learning opportunity in the absence of the target within cue-only repetitions. Interestingly, cue-only repetitions enhanced retrieval from long-term memory in the massed condition, relative to the spaced condition, indicating that the retrieval cue is effective to the extent that information about the target is available.

On each type of errors, a 2×2 ANOVA was conducted (Table A1: Intrusion and No response). Since the interaction was significant for both error types, we report the results of simple effect analyses (Table A2: Intrusion and No response).

Table 1 Mean proportions and latencies (log scale seconds) of intrusions and no responses

			Item type		
			Cue target	Cue only	Control
Proportion					
Experiment 1	Intrusion	Massed	.48 (.04)	.46 (.04)	–
		Spaced	.42 (.04)	.49 (.04)	–
	No response	Massed	.27 (.04)	.30 (.04)	–
		Spaced	.21 (.03)	.31 (.04)	–
Experiment 2	Intrusion	Massed	.51 (.03)	.49 (.03)	.50 (.03)
		Spaced	.41 (.03)	.50 (.03)	.52 (.03)
	No response	Massed	.18 (.03)	.18 (.03)	.21 (.03)
		Spaced	.13 (.02)	.21 (.03)	.19 (.03)
Experiment 3	Intrusion	Spaced	.34 (.03)	–	.38 (.03)
	No response	Spaced	.18 (.02)	–	.23 (.03)
Latency					
Experiment 1	Intrusion	Massed	1.70 (0.06)	1.68 (0.07)	–
		Spaced	1.77 (0.06)	1.63 (0.06)	–
	No response	Massed	1.39 (0.10)	1.34 (0.10)	–
		Spaced	1.45 (0.11)	1.31 (0.09)	–
Experiment 2	Intrusion	Massed	1.60 (0.05)	1.63 (0.06)	1.60 (0.05)
		Spaced	1.60 (0.05)	1.43 (0.04)	1.42 (0.04)
	No response	Massed	1.49 (0.10)	1.47 (0.10)	1.47 (0.10)
		Spaced	1.55 (0.12)	1.34 (0.10)	1.40 (0.09)
Experiment 3	Intrusion	Spaced	1.87 (0.05)	–	1.46 (0.04)
	No response	Spaced	1.92 (0.08)	–	1.38 (0.06)

Note. Standard errors of the mean are in parentheses

For both error types, in the massed condition, there was no difference between cue–target and cue-only repetitions, whereas in the spaced condition, there were fewer errors for cue–target repetitions than for cue-only repetitions. In addition, for cue–target repetitions, there were fewer errors in the spaced condition than in the massed condition, but for cue-only repetitions, there was no difference between the spaced and massed conditions. Overall, an increase in correct

responses (in the spaced cue–target condition) was accompanied by a decrease in both error responses.

Recall latency Figure 2 shows mean correct recall latencies (Fig. 2b). A 2×2 ANOVA showed that recall latency was shorter in the spaced condition than in the massed condition, and that neither the effect of item type nor the interaction was significant (Table A1: Correct). On average, strong items

Fig. 2 Mean proportions (a) and latencies (b) of correct recall as a function of repetition type (massed and spaced) and item type (cue–target and cue only) of Experiment 1. Error bars depict ± 1 standard error of the mean

(spaced-repetition items) appeared to be recalled faster than weak items (massed-repetition items).

However, special attention must be paid to cue-only repetitions across the massed and spaced conditions. For cue-only repetitions, massed-repetition items (strong items) were recalled more slowly than were spaced-repetition items (weak items), as seen in Fig. 2. In addition, despite no significant interaction on recall latency, the pattern of results was reversed in the massed and spaced conditions: It appeared that the items of cue–target repetitions (strong items) were recalled more slowly than those of cue-only repetitions (weak items), but in the spaced condition only. Considering that multiple comparisons increase the Type I error rate, we analyzed recall-latency distributions to further investigate these results and report the data in the next section.

On each type of error latencies, a 2×2 ANOVA was conducted (Table A1: Intrusion and No response). For both error types, the main effect of repetition type was not significant. The main effect of item type was significant, which showed that both error latencies were longer for cue–target repetitions than for cue-only repetitions. For intrusions, the interaction was marginally significant, showing that spaced cue–target repetitions led to longer error latencies, relative to spaced cue-only repetitions, as seen in Table 1 (although the interaction was not significant for response failures, the same pattern of data emerged). This finding is new, and we return to this in the General Discussion.

Recall-latency distributions Figure 3 shows the recall-latency distribution for correct responses, the best fitting ex-Gaussian function obtained using maximum likelihood estimation (Zandbelt, 2014; see further Lacouture & Cousineau, 2008; Lewandowsky & Farrell, 2010), and the three parameters (μ , σ , and τ) for each condition: the massed cue–target and cue-only conditions (Figs. 3a, b, respectively); and the spaced cue–target and cue-only conditions (Figs. 3c, d, respectively).

To determine whether parameter estimates were different across the conditions, we created 10,000 samples with replacement from the observed data set, using a nonparametric bootstrapping technique (Efron & Tibshirani, 1993; e.g., Jang, Wixted, & Huber, 2011), and the data set of each sample was then fit with the ex-Gaussian, exactly as for the observed data set. Figure 4 shows mean parameter estimates of the 10,000 samples for massed (Fig. 4a) and spaced (Fig. 4b) repetitions. For massed repetitions, estimates of μ , σ , and τ were similar between cue–target and cue-only repetitions. For spaced repetitions, estimates of τ were different whereas estimates of μ and σ were similar between cue–target and cue-only repetitions. We calculated the difference in τ between cue–target and cue-only repetitions, as well as differences in μ and σ , and examined whether the differences in each parameter were statistically significant (e.g., Rotello & Zeng, 2008). For example, we subtracted τ in the massed (or spaced) cue-only condition from τ in the massed (or spaced) cue–target condition. We did the

same calculation between the massed and spaced conditions, subtracting τ in the spaced condition from τ in the massed condition for cue–target (or cue-only) repetitions. For all experiments, there were no significant differences in μ and σ . Below, we report the results of τ in detail for each experiment.

Table 2 shows mean differences in τ and 95% confidence intervals (CIs) for each of the four pairwise comparisons. The CIs include zero for massed repetitions, so we cannot reject the null hypothesis that τ is not different between cue–target and cue-only repetitions in the massed condition. In addition, the CIs include zero for cue–target repetitions, indicating that τ is not different between the massed and spaced conditions for cue–target repetitions. By contrast, for spaced repetitions, the CIs do not include zero, indicating that τ is greater for cue–target repetitions than for cue-only repetitions when repetitions were spaced. Finally, for cue-only repetitions, the CIs do not include zero, indicating that τ is greater in the massed condition than in the spaced condition when the cue alone was repeated. In other words, the main effect of repetition type on recall latency as reported above was due to cue-only repetitions.

These latency-distribution results suggest interesting accounts of retrieval processes. First, strong items (in the spaced cue–target condition) were not recalled faster than weak items (in the massed cue–target condition) in a mixed-strength situation. It is assumed that spaced repetitions strengthen the link between items and their contexts, compared with massed repetitions (Malmberg & Shiffrin, 2005). Possibly, in the SAM model, context strength during the sampling process may play an important role in slow and more accurate recall in the spaced cue–target condition. However, the separate storage hypothesis cannot be ruled out in that recall in the spaced cue–target condition was as slow as recall in the massed cue–target condition although recall accuracy was higher in the spaced cue–target condition. Spaced cue–target repetitions can be separately stored, which increases search set size as well as sampling strength. Second, strong items (in the spaced cue–target condition) were recalled more slowly than weak items (in the spaced cue-only condition). It is somewhat surprising, but it is possible that participants engaged in covert retrieval attempts given the cue when information about the target was no longer available in short-term memory. If such a form of retrieval practice is successful, then that can help to access information concerning the target and speed-up the retrieval process while not necessarily increasing accuracy (Jang et al., *in press*). The next two experiments investigated these possibilities.

Experiment 2

In Experiment 1, we found slow and more accurate recall in a mixed list when manipulating repetition type and item type for repetition. This finding, posing a problem for the relative strength model, has not been reported, and Experiment 2

Fig. 3 Recall-latency distributions for correct responses of Experiment 1. Solid curves represent the best fitting ex-Gaussian. The parameters of the ex-Gaussian are shown

was designed to further examine recall latency for strong and weak items in the same mixed-strength situation of Experiment 1, but with two additions.

First, we included a pretest trial before the repetition manipulation occurred for each condition. Then, each item was categorized as a retrievable or nonretrievable item, depending on

Fig. 4 Mean ex-Gaussian parameter values of Experiment 1. Parameters were estimated using a nonparametric bootstrapping technique with 10,000 samples. Error bars depict ± 1 standard deviation

Table 2 Mean differences and confidence intervals (CIs) for the difference in τ of Experiment 1

		95% CI lower bound	Mean	95% CI upper bound
Massed	Cue–target – Cue only	–426	105	645
Spaced	Cue–target – Cue only	169	872	1,786
Cue–target	Massed – Spaced	–996	–33	755
Cue only	Massed – Spaced	309	734	1,173

Note. CIs were estimated with a nonparametric bootstrap technique (10,000 samples). Bold type indicates the CIs do not include zero, or the null hypothesis is rejected

response outcome on the pretest: An item that was or was not retrieved on the pretest, respectively. Comparing restudy and test practice effects, Jang et al. (2012) found for an immediate test, as in the present study that for retrievable items, there was a very small benefit of restudy, whereas for nonretrievable items there was a large benefit of restudy. In like manner, we considered the role of retrievability in the present experiment, where cue–target repetitions were equivalent to restudy trials in Jang et al.’s study, to better understand slow retrieval process for strong items in the spaced cue–target condition, as found in the first experiment. To examine performance for retrievable and nonretrievable items separately, more participants were recruited. It should be noted that the procedure of the pretest was different from that of the initial test for test practice, as found in the literature of retrieval practice, in that the initial test typically took place during the practice phase once all items were studied (e.g., the study and test practice phases followed by the final test phase). In the present experiment, the pretest was presented during the study phase (i.e., the study phase including pretest trials, followed by the test phase), considering the massed and spaced conditions.

Second, we included the control condition in which neither cue–target pair nor cue alone was repeated during the study phase (i.e., no repetition). In this manner, the control condition served as the pure test practice condition, compared with the two experimental conditions of item type, which consisted of cue–target or cue-only repetitions as well as the pretest (or test practice). If faster (and less accurate) recall resulted from covert retrieval practice in the cue-only condition, then there would be no or little difference in performance between the cue-only and control conditions (compared with performance between the cue–target and control conditions), in that the control condition involved pure, overt retrieval practice by virtue of the pretest. Considering repetition type, there were two control conditions in which the study and pretest trials were massed and spaced, respectively.

Method

Participants Ninety-five undergraduate students at the University of Montana were recruited and received credit for psychology courses in return for their participation.

Design A 2 (repetition type: massed and spaced) $\times$ 3 (item type: cue–target, cue only, and control [no repetition]) within-subjects factorial design was used.

Materials The same word pairs from Experiment 1 and additional 24 pairs (selected from the same word pool) were used in Experiment 2 (i.e., 12 pairs for each condition as in Experiment 1).

Procedure The procedure was identical to that of Experiment 1, except as noted. In Experiment 2, a pretest trial was inserted right before the repetition manipulation occurred. For example, consider two pairs of “A-B” and “C-D” (in the massed and spaced conditions, respectively, as illustrated in Fig. 1). For massed repetitions, after “A-B” appeared on the screen as a study item, participants were asked to type in the target (“B” in this case) when given the cue of “A- ” (i.e., pretest). Then, either “A-B” (in the cue–target condition) or “A- ” (in the cue-only condition) was presented, or no repetition occurred (in the control condition). The next study item was then presented. For spaced repetitions, once all 12 pairs per block were presented, a pretest trial (“C- ”) was given. After participants typed in their response, either “C-D” (in the cue–target condition) or “C- ” (in the cue-only condition) was presented, or no repetition occurred (in the control condition). Then, the next pretest trial was presented.

Results and discussion

The complete results from statistical analyses of Experiment 2 are reported in Appendix B (Tables B1 and B2).

Recall response probability Figure 5 shows mean proportions correct recall (Fig. 5a). A 2×3 ANOVA was conducted (Table B1: Correct) with repetition type (massed and spaced) and item type (cue–target, cue only, and control). The main effects and interaction were significant. Recall accuracy was higher in the spaced condition than in the massed condition. Recall accuracy was higher for cue–target repetitions than for cue-only repetitions and for the items in the control condition, with Bonferroni corrections. Simple effect analyses (Table B2: Correct) showed that for spaced repetitions, recall accuracy

was higher in the cue–target condition than in the cue-only condition and in the control condition, with Bonferroni corrections, and there was no difference between the cue-only and control conditions. Recall accuracy was not different between the three massed conditions. In addition, for cue–target repetitions, recall accuracy was higher in the spaced condition than in the massed condition. However, the pattern was in the opposite direction for cue-only repetitions: Recall accuracy was higher in the massed condition than in the spaced condition. For the items in the control condition, there was no difference between the two conditions. These findings replicated those of Experiment 1, suggesting that to improve the accuracy of long-term memory, cue–target (not cue-only) repetitions should be spaced.

Next, we examined performance on the pretest and conducted conditional analyses. Performance on the pretest is reported in Table 3. As expected, recall accuracy reaches a ceiling for massed repetitions and shows equality between the three conditions for each repetition type because the

pretest was given before the manipulation occurred. For conditional analyses, performance was broken down as a function of response outcome on the pretest. First, not surprisingly, initially incorrect responses, or nonretrievable items, were mostly incorrect in all conditions except for the spaced cue–target condition ($M = .34$)—further analysis was impossible, indicating that additional spaced study opportunity is crucial to correcting initial errors. Second, Fig. 5c shows conditional probabilities for retrievable items (conditioned on correct pretest responses). A 2×3 ANOVA revealed (Table B1: Correct) that recall accuracy was higher in the spaced condition than in the massed condition. This is not surprising, considering that massed repetitions yielded ceiling effects on the pretest. The main effect of item type was not significant, indicating that for retrievable items there is no benefit of item repetition. This is a different but not discordant result from that of Jang et al. (2012), considering not only that the procedure was different but also that in their experiment, the study benefit for retrievable items was fairly small (.03 increase in the proportion of

Fig. 5 Mean proportions (**a** and **c**) and latencies (**b** and **d**) of correct recall as a function of repetition type (massed and spaced) and item type (cue–target, cue only, and control [no repetition]) of Experiment 2. **a–b** All correct responses regardless of response outcome on the pretest. **c**

Conditional probabilities for retrievable items (conditioned on correct pretest responses). **d** Correct recall latencies for retrievable items. Error bars depict ± 1 standard error of the mean. Note the y-axis scale is different between panels **a** and **c**

Table 3 Mean proportions and latencies (log scale seconds) of correct recall on the pretest of Experiments 2 and 3

			Item type		
			Cue target	Cue only	Control
Accuracy	Experiment 2	Massed	.98 (.005)	.98 (.004)	.98 (.004)
		Spaced	.29 (.025)	.30 (.026)	.29 (.026)
	Experiment 3	Spaced	.99 (.002)	–	.40 (.022)
		Spaced	.99 (.002)	–	.40 (.022)
Latency	Experiment 2	Massed	0.96 (0.05)	0.93 (0.05)	0.96 (0.05)
		Spaced	1.40 (0.04)	1.42 (0.05)	1.39 (0.05)
	Experiment 3	Spaced	0.59 (0.06)	–	1.45 (0.03)
		Spaced	0.59 (0.06)	–	1.45 (0.03)

Note. Standard errors of the mean are in parentheses

items recalled) for the immediate test. The interaction was not significant.

Finally, on each type of errors, a 2×3 ANOVA was conducted (Table B1: Intrusion and No response). Similar to the results of Experiment 1, the interaction effect exists for both error types, and simple effect analyses (Table B2: Intrusion and No response) showed a consistent pattern of results that cue–target repetitions in the spaced condition led to fewer errors (as well as more correct responses), relative to the rest of the conditions for both error types.

Recall latency Figure 5 shows mean correct recall latencies for all correct responses regardless of response outcome on the pretest (Fig. 5b) and those for retrievable items (Fig. 5d). First, a 2×3 ANOVA was conducted for all correct responses (Table B1: Correct). The main effect of repetition type was significant, which showed that recall latency was shorter in the spaced condition than in the massed condition. Consistent with the finding of the first experiment, however, for cue-only repetitions, recall latency was longer for massed-repetition items (strong items) than for spaced-repetition items (weak items), as seen in Fig. 5b. In addition, for the items in the control condition, recall latency was longer in the massed condition than in the spaced condition even when there was no difference in accuracy. We looked into these data in greater detail, analyzing recall-latency distributions (to be reported in the next section).

There was a main effect of item type, which showed that recall latency was longer for cue–target repetitions (strong items) than for cue-only repetitions and for the items in the control condition (weak items), with Bonferroni corrections. The interaction was not significant. Interestingly, however, the same pattern of data emerged in recall accuracy (Fig. 5a) and latency (Fig. 5b) across the three spaced conditions: higher accuracy and longer latency in the spaced cue–target condition. We further analyzed the latency data in these spaced conditions (Table B1: Correct). With Bonferroni corrections, recall latency was longer in the cue–target condition than in

the cue-only and control conditions, and there was no difference between the cue-only and control conditions. For completeness, there was no difference between the three massed conditions (Table B1: Correct). In other words, the main effect of item type as reported above was due to spaced repetitions.

Next, a 2×3 ANOVA was conducted for retrievable items (Table B1: Correct). The main effect of repetition type was significant, which showed that recall latency was shorter in the spaced condition than in the massed condition. Neither the effect of item type nor the interaction was significant. Specifically, the difference observed in the three spaced conditions for all correct responses (i.e., longer latency for cue–target repetitions) disappeared when only the items correctly recalled on the pretest were analyzed. These indicate that item repetition does not change recall latency for retrievable items.

Finally, on each type of errors, a 2×3 ANOVA was conducted (Table B1: Intrusion and No response). The interaction effect exists for intrusions and marginally for response failures. As seen in the results of simple effect analyses (Table B2: Intrusion) and Table 1, both error latencies were longer in the cue–target condition than in the cue-only and control conditions, but for spaced repetitions only, and there was no difference between the cue-only and control conditions. These results are in agreement with those of Experiment 1, and once again, the consistent finding is new, and we return to this in the General Discussion.

Recall-latency distributions Figure 6 shows the recall-latency distribution for correct responses, the best fitting ex-Gaussian function, and the three parameters for each of the six conditions (Figs. 6a–f). The data in the spaced cue–target condition (Fig. 6d) were split into the two data sets for retrievable and nonretrievable items, based on response outcome on the pretest, and the results are shown in Figs. 6g and h, respectively.

To determine whether parameter estimates were different across the conditions, we created 10,000 samples from the bootstrap method and analyzed the data described for Experiment 1. Figure 7 shows mean parameter estimates of the 10,000 samples for massed (Fig. 7a) and spaced (Fig. 7b) repetitions. Consistent with the data from Experiment 1, for massed repetitions, estimates of μ , σ , and τ were similar across the three conditions. For spaced repetitions, estimates of μ and σ were similar across the three conditions, but estimates of τ were different. Figure 7c shows mean parameter estimates for retrievable items in the three spaced conditions and for nonretrievable items in the spaced cue–target condition. We calculated the difference in τ for 15 pairwise comparisons and CIs, as shown in Table 4: (1) three comparisons among the three massed conditions (i.e., between the cue–target and cue-only conditions; between the cue–target and control conditions; and between the cue-only and control conditions); (2) three comparisons among the three spaced conditions; (3) three comparisons between massed and spaced repetitions

Fig. 6 Recall-latency distributions for correct responses of Experiment 2. **a–f** All correct responses regardless of response outcome on the pretest. **g–h** Correct responses for retrievable and nonretrievable items,

respectively, in the spaced cue–target condition (i.e., the data of panel **d** were split into these two data sets). Solid curves represent the best fitting ex-Gaussian. Parameters of the ex-Gaussian are shown

for each of the cue–target, cue-only, and control conditions; (4) three comparisons among the three spaced conditions for retrievable items; and (5) three comparisons among the three spaced conditions for retrievable items and the spaced cue–target condition for nonretrievable items.

Regarding all items for correct responses (the first nine pairwise comparisons), the CIs include zero for all three comparisons for massed repetitions (the first to third pairwise comparisons); for the comparison between the spaced cue-only

and control conditions (the sixth pairwise comparison); and for the comparison between the massed and spaced conditions for cue–target repetitions (the seventh pairwise comparison). For each of the five, we cannot reject the null hypothesis, which indicates τ is not different between the conditions.

However, the CIs do not include zero for the comparison between the spaced cue–target and cue-only conditions (the fourth pairwise comparison); and for the comparison between the spaced cue–target and control conditions (the fifth

Fig. 7 Mean ex-Gaussian parameter values of Experiment 2. **a–b** All correct responses regardless of response outcome on the pretest. **c** Correct responses for retrievable items in the spaced cue–target, cue-only, and control conditions, and correct responses for nonretrievable items in

the spaced cue–target condition. Parameters were estimated using a non-parametric bootstrapping technique with 10,000 samples. Error bars depict ± 1 standard deviation. Note the y-axis scale is different between panels **b** and **c**

pairwise comparison). These indicate that τ is greater in the spaced cue–target condition, relative to the remaining two spaced conditions. In addition, the CIs do not include zero for the comparison between the massed and spaced conditions for cue-only repetitions (the eighth pairwise comparison), and for the items in the control condition (the ninth pairwise comparison). These indicate that τ is greater in the massed condition than in the spaced condition for each of the cue-only and control conditions. These findings replicate those of Experiment 1, showing that in a mixed-strength situation, strong items (in the spaced cue–target condition) were not recalled faster than weak items (in the massed cue–target condition), and they were recalled more slowly than weak items (in the spaced cue-only and control conditions). In addition, the results extend those of Experiment 1, showing that recall latency remained unchanged between the spaced cue-only and

control conditions. Considering that the control condition is equivalent to the test practice condition in the literature of retrieval practice, this finding indicates that spaced cue-only repetitions involve covert retrieval practice, which leads to faster recall (although recall accuracy was lower compared with spaced cue–target repetitions).

For retrievable items in the three spaced conditions (the 10th to 12th pairwise comparisons), the CIs include zero for each of the three comparisons, indicating no difference in τ . By contrast, the CIs do not include zero for the remaining three comparisons (the 13th to 15th pairwise comparisons) that the spaced cue–target condition for nonretrievable items was compared with another condition for retrievable items, indicating an increase in τ for nonretrievable items. In other words, the increase in τ observed in the spaced cue–target condition was due to nonretrievable items. These indicate that if the items that

Table 4 Mean differences and confidence intervals (CIs) for the difference in τ of Experiment 2

		95% CI lower bound	Mean	95% CI upper bound
Massed	Cue–target – Cue only	–575	–134	272
	Cue–target – Control	–532	–159	222
	Cue only – Control	–414	–25	405
Spaced	Cue–target – Cue only	477	730	985
	Cue–target – Control	106	442	774
	Cue only – Control	–602	–288	14
Cue–target	Massed – Spaced	–323	14	362
Cue only	Massed – Spaced	546	878	1,261
Control	Massed – Spaced	237	615	990
Spaced	Cue–target (Retrievable) – Cue only (Retrievable)	–46	216	483
	Cue–target (Retrievable) – Control (Retrievable)	–261	76	407
	Cue only (Retrievable) – Control (Retrievable)	–438	–140	142
Spaced	Cue–target (Retrievable) – Cue–target (Nonretrievable)	–1,711	–1276	–859
	Cue only (Retrievable) – Cue–target (Nonretrievable)	–1,898	–1492	–1,116
	Control (Retrievable) – Cue–target (Nonretrievable)	–1,806	–1352	–909

Note. CIs were estimated with a nonparametric bootstrap technique (10,000 samples). Bold type indicates the CIs do not include zero, or the null hypothesis is rejected

were not recalled after the first presentation are repeated some time later from the study trial, they are relatively likely to be recalled, which increases overall accuracy, and at the same time, it takes longer to recall them, which increases overall latency. In this sense, it is possible that a second presentation of a nonretrievable item creates another copy of the item (Raaijmakers, 2003), and it may therefore take longer to retrieve these relatively new copies of traces because of continued resampling of already recalled items (retrievable items).

Experiment 3

Experiment 3 was designed to assess recall latency for strong and weak items in pure lists as compared with the two prior experiments in which a mixed list was used. According to the relative strength model, recall latency should remain unchanged whether all of the items are strong or all of the items are weak because the relative strength of a strong item is the same as that of a weak item. However, previous research found that when repetitions were spaced, strong items were recalled more slowly than weak items in pure lists (Bousfield et al., 1954; Unsworth, 2015). This may be due to the possibility that when the memory traces are strengthened by repetition, all of the competitors are not equally strengthened, and some items may be still nonretrievable. To investigate this possibility, in the present experiment, we considered spaced

repetitions in which cue–target and control conditions were compared and additionally included the pretest similar to that of Experiment 2, but the pretest took place after the repetition manipulation. This change allowed us to examine the latency effect of item repetition particularly for retrievable items when they were strengthened by the second presentation. Specifically, we investigated whether greater strength by item repetition leads to slower recall in pure lists even when the items have been retrieved. In this manner, the cue–target condition is in fact equivalent to the restudy and test practice condition (for strong items), whereas the control condition is equivalent to the pure test practice condition (for weak items).

Method

Participants One hundred and eighty undergraduate students at the University of Montana were recruited and received credit for psychology courses in return for their participation. They were randomly assigned to one of two conditions (to be described in the Design section), and, consequently, each condition consisted of 90 participants.

Design The design was a between-subject design composed of cue–target and control (no repetition) conditions.

Materials The 60 word pairs from Experiment 1 were used in Experiment 3 (i.e., 48 pairs included as experimental trials for

data analysis). Each of the two conditions consisted of two study blocks, and for each block, 24 pairs were randomly assigned.

Procedure The procedure was similar to that of the spaced cue–target and control conditions in Experiment 2 except as noted. In Experiment 3, a pretest trial was inserted right after the manipulation occurred. For example, consider two pairs of “A–B” and “C–D” (as illustrated in Fig. 1). In the cue–target condition, once all 24 pairs per block were presented, the cue–target pair (“A–B”) was repeated for 4 s, followed by the pretest trial (“A–___”), in which participants were asked to type in the target (“B” in this case). The next cue–target pair (“C–D”) was then repeated. In the control condition, once all 24 pairs per block were presented, participants took the pretest for the items assigned to the condition one at a time: e.g., “A–___” and then “C–___” for the two studied items.

Results and discussion

The complete results from statistical analyses of Experiment 3 are reported in Appendix C.

Recall-response probability Performance on the pretest is reported in Table 3. Figure 8 shows mean proportions correct recall (bars). For all correct responses, regardless of response outcome on the pretest, recall accuracy was higher in the cue–target condition than in the control condition as shown in Fig. 8a. The key analysis of interest considered performance for retrievable items, and conditional analyses revealed that recall accuracy was higher in the control condition than in the cue–target condition, as shown in Fig. 8b. This is not surprising in that most pretest responses were correct in the cue–target condition. Finally, for both error types (Intrusion and No response), the difference between the two conditions did not reach statistical significance.

Recall latency Figure 8 shows mean correct recall latencies (dotted lines) for all correct responses regardless of response outcome on the pretest (Fig. 8a) and for retrievable items (Fig. 8b). Both correct recall latencies were longer in the cue–target condition than in the control condition: Recall latency was still slower for strong items even if they were retrieved on the pretest. Finally, error latencies (Intrusion and No responses) were longer in the cue–target condition than in the control condition, consistent with the findings of Experiments 1 and 2 (all error latency results will be discussed in the General Discussion).

Recall-latency distributions Figure 9 shows the recall-latency distribution for correct responses, the best fitting ex-Gaussian function, and the three parameters for the cue–target (Fig. 9a) and control (Fig. 9b) conditions. The remaining two panels (Figs. 9c–d) show the results for retrievable items for each condition. Figure 10 shows mean parameter estimates (Fig. 10a) of the 10,000 samples, which were obtained from the bootstrap method described for Experiment 1. The results for retrievable items are shown in Fig. 10b. Estimates of μ and σ were similar across the conditions, but estimates of τ were different, as seen in each panel. Specifically, the difference in τ was calculated for the data in each panel. Mean difference values as well as CIs are reported in Table 5. The CIs do not include zero not only for all items but also for retrievable items, indicating that τ is greater in the cue–target condition than in the control condition for each. These results replicate and extend those of prior research (Bousfield et al., 1954; Unsworth, 2015). If an item received an additional spaced learning opportunity, recall latency was longer for the item than for another item that was not repeated, even if both items were later recalled. This suggests that spaced repetitions are stored separately with variability for each copy of the items in one’s memory.

Fig. 8 Mean proportions (bars) and latencies (dotted lines) of correct recall as a function of item type (cue–target and control [no repetition]) of Experiment 3. **a** All correct responses regardless of response outcome on the pretest. **b** Conditional probabilities for retrievable items

(conditioned on correct pretest responses) and correct recall latencies for retrievable items. Error bars depict ± 1 standard error of the mean. Note the primary y-axis scale is different between the panels

Fig. 9 Recall-latency distributions for correct responses of Experiment 3. **a–b** All correct responses regardless of response outcome on the pretest **c–d** Correct responses for retrievable items. Solid curves represent the best fitting ex-Gaussian. Parameters of the ex-Gaussian are shown

General discussion

We investigated how item repetition affects recall accuracy and latency examining recall-latency distributions in a cued-recall paradigm. If recall latency reflects solely relative strength assumed by the relative strength model, then strong items will be recalled faster than weak items in mixed lists, and there will be no change in recall latency in pure lists. In three experiments of the present study, item repetition was used as a means of strengthening, and the latency-distribution analysis yielded the following results. First, recall latency did not differ when cue–target repetitions were spaced

versus massed in a mixed list, although accuracy increased with spaced repetitions. Second, recall latency was longer in a mixed-strength situation when cue–target repetitions were spaced, relative to cue-only repetitions, even if cue–target repetitions led to higher accuracy. Third, slow recall with spaced cue–target repetitions in the mixed list was due to nonretrievable items, or the items that were not retrieved on a pretest. Fourth, in pure lists, spaced cue–target repetitions increased both accuracy and latency, although these strong items were retrieved after repetitions. These findings pose a challenge for relative-strength explanations of latency. Finally, spaced cue–target repetitions also resulted in slow incorrect

Fig. 10 Mean ex-Gaussian parameter values of Experiment 3. **a** All correct responses regardless of response outcome on the pretest. **b** Correct responses for retrievable items. Parameters were estimated using a

nonparametric bootstrapping technique with 10,000 samples for each condition. Error bars depict ± 1 standard deviation

recall for both intrusions and response failures (as new findings in all experiments).

To shed light on how to explain these results, we consider the general framework of SAM, including its possible variants, and different but related models. It should be emphasized that the purpose of the present study was not to compare the relative strength model to other models, such as SAM, and that a completely different approach, rather than SAM, could provide better answers, or even new models should be developed to account for all the results. However, we focus mainly on the SAM approach, first because it considers not only item strengths but also interitem associations and context strengths (which are not distinguished in the relative strength model), and such features and assumptions have been successfully able to account for various memory phenomena. Secondly, and more importantly, the findings of the present study indeed bear on the critical assumptions and features of the SAM model.

Sampling, context, and recall latency

We begin with the consideration of sampling. It is informative to compare key assumptions between the original SAM model (Gillund & Shiffrin, 1984; Raaijmakers & Shiffrin, 1980, 1981) and Shiffrin et al.'s (1990) SAM, which includes additional assumptions to account for the list-strength effect—performance decreases as the strengths of other items increase. It is not only because the difference in the assumptions

between the two is important to predict the list-strength effect but because it also helps to explain our results of cue–target repetitions, or slow and more accurate recall when cue–target repetitions were spaced. The original SAM model posits that individual items are stored separately in memory when a new item (as in a list-length manipulation) or a repetition of an old item (as in a list-strength manipulation) is presented. The core of the separate storage hypothesis is how item repetitions are stored. If item repetitions are separately stored, a positive list-strength effect is predicted because each separately stored item repetition adds its own activation to the summed activations of all memories, and this increases the noise in the retrieval process and thereby decreases performance.

However, list-strength results are not always positive (e.g., Ratcliff et al., 1990; also see Wilson & Criss, 2017, for recent replications in cued recall). As a solution, Shiffrin et al. (1990) proposed a variant of SAM with a differentiation assumption. Specifically, the similarity between a stored item and an item cue decreases as the stored item becomes strong. The more dissimilar the item cue and the stored item, the less the stored item is activated by the item cue. At the same time, the context cue will produce more activation of strong items as strong-item repetitions are accumulated. For strong items, this increase in response to the context cue is offset by a tendency to be less activated in response to the item cue. Strong and weak items thereby produce the same net level of activation (leading to a null list-length effect). The differentiation hypothesis predicts

Table 5 Mean differences and confidence intervals (CIs) for the difference in τ of Experiment 3

	95% CI lower bound	Mean	95% CI upper bound
Cue–target – Control	633	774	916
Cue–target (Retrievable) – Control (Retrievable)	684	823	965

Note. CIs were estimated with a nonparametric bootstrap technique (10,000 samples). Bold type indicates the CIs do not include zero, or the null hypothesis is rejected

positive to negative list-strength effects by using context and item cues (also see Murnane & Shiffrin, 1991), based on the assumption that repetitions are accumulated into a single trace.

It appears that the separate storage hypothesis, relative to the accumulated storage hypothesis, can handle the finding that spaced cue–target repetitions increased both accuracy and latency (particularly, in pure lists) without additional assumptions. As suggested by Unsworth (2015), if item repetitions are separately stored, these increase the number of copies of an item in a search set. That is, each repetition creates a new copy, which increases not only sampling strength of the item but also the size of the search set. Since multiple copies are stored in memory, finding the item will be slow given a larger search set. We additionally suggest that the core of the separate storage hypothesis is whether new context information also contributes to an increase in the size of the search set as new copies are created. It is because without consideration of context strengthening, the separate storage hypothesis may provide qualitatively the same prediction as the accumulated storage hypothesis (e.g., Wixted et al., 1997).

Regarding item repetition, the importance of context information has been demonstrated by Malmberg and Shiffrin (2005), in which a version of SAM/REM was applied. According to their “one-shot” hypothesis, spaced repetitions strengthen the link between items and their contexts, whereas other strengthening manipulations (e.g., increasing study time) affect only the item strength. Each time an item is presented, one shot of context is stored in a trace. During massed repetitions (or during extra study time beyond sufficient study time), no additional shots of context are stored, and so only the item strength is increased by massed repetitions. However, for spaced repetitions, both the item and context information are incremented. It is worth stressing that context information rises to a maximum value during sampling to account for the result that in the mixed list, recall latency was not faster for spaced cue–target repetitions than for massed cue–target repetitions despite the accuracy advantage for spaced-repetition items. Possibly, compared to massed repetitions, spaced repetitions increase the likelihood of sampling the items due to an increase in context information, leading to higher accuracy. At the same time, however, because each trace more includes its own variability, it may affect the speed of search on all the traces of an item within a search set, and thereby there is no latency advantage for strong items with spaced repetitions. Nonetheless, the details of such possibilities should await specific implementations of the model.

The separate storage hypothesis is certainly open to debate. It provides a solution to predict latency effects, involving a sampling process having properties like SAM. However, the accumulated storage assumption holds for a large class of models, including SAM and REM, predicting accuracy effects, including the list-strength effect. Murnane and Shiffrin (1991) suggested that when an item is repeated in a similar context, the

new information is typically accumulated in the old memory, so that repetitions are updated into one trace, but a separate trace can also be stored when the contexts are substantially different. Along the same line, Shiffrin (2003) wrote, “The traces are stored separately, although some are formed through accumulation of multiple events. . . . When an event is similar to one or more earlier events (e.g., a nominal repetition is taking place) there is a tendency to store a new episodic trace, but also a tendency to store additional information in the previously stored trace” (p. 343). In addition, Raaijmakers (2003) reached the same conclusion, developing a new model with the SAM approach. His model assumes that when an item is repeated, if it is retrieved, then repetition strengthens the memory for the item, and if not, then it creates a new trace, which is separately stored in memory. Our results are supportive of the account, showing that slow recall with spaced cue–target repetitions was due to nonretrievable items. Probably, if an item is retrieved on a pre-test, additional copies of the item is not necessary, and its trace is updated in memory. This does not affect the size of the search set, and recall latency thereby is not influenced. By contrast, if an item is not initially retrieved but is later repeated during study, a new memory trace is stored, leading to a larger search set, and so recall latency becomes longer.

One could argue that our findings in cued recall and theoretical accounts of accuracy in cued recall or recognition have nothing to do with the relative strength model, because the model offers an account of latency in free recall. For example, they might point out that the list-strength effect is task dependent—typically, positive list-strength effects in free recall, but null to negative list-strength effects in cued recall and recognition. However, prior research failed to find support for the relative strength model in a free-recall paradigm (e.g., Bousfield et al., 1954; Hopper & Huber, 2018a; Unsworth, 2015), and the same pattern of data observed in free recall, inconsistent to the relative strength model, was also reported in a cued-recall paradigm (e.g., Hopper & Huber, 2018a; van den Broek et al., 2014). Somewhat different assumptions are indeed needed to account for these findings. Moreover, distinguishing two primary sources of information in memory, or item and context, assumed in SAM and REM, we take a stab at explaining the findings in the present study as well as previous studies, and the importance of context information we considered is widely assumed, particularly in a free-recall paradigm. The point of this discussion is not to evaluate the relative strength model but to raise the possibility to provide a unified account of both accuracy and latency effects, rather than task-specific explanations of either accuracy or latency.

Recovery of items and recall latency

Next, we consider the possibility that some forms of learning selectively affect the recovery process, leading to a change in recall latency. Learning occurs not only from studying

information repeatedly, such as spaced repetitions, but also from retrieving information. The benefits of restudy and test practice typically appear in terms of accuracy on an immediate test and on a delayed test, respectively. Kornell, Bjork, and Garcia (2011) suggested that while restudy strengthens all items, test practice without feedback produces bifurcated item distributions, with stronger items after successful retrieval and with weak items after retrieval failure. When a test takes place immediately, the benefit of restudy appears because all repeated items receive increases in memory strength, whereas only some retrieved items receive increases in memory strength after test practice. However, the bifurcated item distributions after test practice make retrieved items appear to be forgotten more slowly than are restudied items, and so the benefit of test practice appears when a test takes place after a delay. In this sense, the bifurcation model can describe the crossover interaction between practice type and retention interval typically observed in the literature of retrieval practice (for a review, see Roediger & Karpicke, 2006) and the results of the present study for the immediate test: When repetitions were spaced, cue–target repetitions increased recall accuracy (and latency), compared with cue-only repetitions. Considering the bifurcation model, it may be because some items in the cue-only condition become even stronger than the items strengthened by cue–target repetitions (assuming cue-only repetitions involve covert retrieval practice that is as effective as overt retrieval practice). However, the bifurcation model assumes that successful retrieval during test practice provides a large increase in memory strength without explaining why test practice should be more effective than restudy. Moreover, assuming that forgetting reflects decreases in memory strength, the bifurcation model cannot explain various types of interference phenomena linked to memory loss.

Incorporating interference results, Huber, Tomlinson, Jang, and Hopper (2015) suggested that learning in the recovery process can be distinguished from the learning that underlies the sampling process. They proposed an extension to the SAM model that if a sampled memory involves recovery for more than one item, then there will be item-specific learning effects during the recovery process, and so successful recall during test practice leads to a selective strengthening of the recovery process. Specifically, their SAM-RI (SAM with recovery interference) model assumes that study practice enhances sampling strengths of the studied items, whereas test practice does the same but also enhances the recovery strength of successfully recalled items. For example, consider a pair as a study item, “A–B”. Successful recall of “B” given the cue of “A” will increase the association between the cue and the target. The overt (and somewhat covert, as well in the present study) production of “B” will make the memory resistant to forgetting, because it becomes likely that the memory will be recovered regardless of how many times that the target item is sampled.

To account for learning from retrieval practice for latency data, Hopper and Huber (2018a) proposed a model termed the PCR (primary and convergent retrieval memory) model that still holds the idea of the sampling and recovery processes. The PCR model assumes that some forms of learning (e.g., repeated study practice) strengthen associations between retrieval cues and the target memory, supporting the sufficiency of memory, whereas other forms of learning (e.g., retrieval practice) strengthen associations between the features that comprise a target memory, resulting in fewer decoding steps and faster retrieval. According to the relative strength model, retrieval latency reflects solely the sampling of items (specifically, item content, itself) from a search set. However, the PCR model suggested that recall latency reflects both a relative competition (i.e., the relative strength of an item in a search set) and the time needed to recover the item, which is determined by absolute strength. If any forms of retrieval practice increase the absolute strength of recalled items and speed up recovery, this explains shorter latencies after the pretest with no additional restudy and for cue-only repetitions observed in the present study.

However, Rohrer (1996) argued that the absolute strength of an item (recovery) determines whether it is recalled, whereas the relative strength of an item (sampling) determines when it is recalled. According to his relative strength with threshold model, items are sampled in a search set, based on the relative strength rule, until all items are found, and then a sampled item is recovered into consciousness only when its absolute strength exceeds a fixed threshold. That is, all items can be sampled, but only those items whose strength exceeds the threshold can be recalled. Although Rohrer showed that recall latency was proportional to list length but independent of recall total, replicating his previous findings (Rohrer & Wixted, 1994), it is unclear how to explain the findings of shorter latencies for weak items in a mixed-strength situation. Future research should be directed at examining whether the distinction between sampling and recovery provides the key for explaining the observed dissociations between latency and accuracy.

Finally, the results of the three experiments yielded slow incorrect recall (as well as slow correct recall) following spaced cue–target repetitions even if they increased recall accuracy. This is somewhat interesting because most models, including the SAM and relative strength models, provide specific predictions regarding only correct responses, not error responses. However, Hopper and Huber (2018b) recently found that after a failure to recall during test practice, failure to recall on a subsequent test was also faster, compared with restudy. According to their PCR model, it is because interitem associations are learned even after retrieval failure (i.e., learning from the failure to recall), as the learning rule of the model concerns the temporal activation of features. That is, a dynamic recovery process leads to not only faster correct recall but

also faster recall failure. It should be noted that Hopper and Huber measured failure latency, rather than error latency, which is the time taken to give a “can’t recall” response (for correct recall, a “recall” response). Our study measured error latencies of intrusions and response failures, separately. Despite the difference in measurements, the pattern of results was the same in both studies, suggesting that the account of the PCR model is generalized: The learning of associations between the features of an item becomes gradually active over retrieval attempts.

Concluding remarks

The present study provides evidence that recall accuracy and latency depend not only on whether a relative competition between items in a search set, possibly including multiple copies of some items after spaced repetitions, but also on how information about the target of an item is experienced and practiced during spaced repetitions of the item. We cast some light on how to explain slow and more accurate recall when item repetitions are spaced, incorporating the assumption that item and context information strengthens the traces of an item and preserving the nature of sampling and recovery of

the SAM model. The distinction between sampling and recovery made it possible not only to explain various phenomena of memory but also to quantify assumed underlying processes of memory. The findings of the present study point out the importance of investigating recall latency as well as accuracy as a means of better understanding search and retrieval processes. Starting with the Atkinson and Shiffrin’s (1968) framework, the successive models, such as SAM, have been quite fruitful. Developing a REM-based model for cued recall (and recognition), Diller et al. (2001) noted that the relationship between recall accuracy and latency is “a very complex matter” (p. 415). A successor requires quantitative tests to handle the specifics of recall accuracy and latency in various experimental paradigms.

Author note Yoonhee Jang, Department of Psychology, University of Montana; Heungchul Lee, Net Intelligence & Research, Seoul, South Korea

We thank David E. Huber for many helpful discussions concerning this study, and three anonymous reviewers and Richard M. Shiffrin for valuable comments on an earlier version of this article.

Appendix A

The complete results from analyses of variance (ANOVAs) of Experiment 1 are reported in Table A1. The results from follow-up simple effect tests of the interactions are reported in Table A2.

Table 6 Complete ANOVAs of Experiment 1

	Correct			Intrusion			No response		
	<i>F</i>	<i>p</i>	η_p^2	<i>F</i>	<i>p</i>	η_p^2	<i>F</i>	<i>p</i>	η_p^2
Proportion									
R	5.78	.019	.08	<1			3.37	.071	
I	49.47	<.001	.43	3.75	.057		27.06	<.001	.29
R × I	38.57	<.001	.37	15.03	<.001	.19	9.79	.003	.13
Latency									
R	24.84	<.001	.38	<1			<1		
I	<1			5.39	.024	.10	5.86	.022	.16
R × I	1.20	.280		3.83	.056		<1		

Note. Effect size (η_p^2) is reported only when the *F* value was significant. R = repetition type (massed and spaced); I = item type (cue–target and cue only)

Table 7 Simple effect tests following the interactions between repetition and item type of Experiment 1

	<i>t</i>	<i>p</i>	<i>d</i>
Proportion: Correct			
Massed: cue–target vs. cue only	0.48	.64	
Spaced: cue–target vs. cue only	8.42	<.001	0.79
Cue–target: massed vs. spaced	−5.84	<.001	0.56
Cue only: massed vs. spaced	1.94	.057	
Proportion: Intrusion			
Massed: cue–target vs. cue only	1.20	.233	
Spaced: cue–target vs. cue only	−3.63	.001	0.24
Cue–target: massed vs. spaced	3.10	.003	0.20
Cue only: massed vs. spaced	−1.66	.103	
Proportion: No response			
Massed: cue–target vs. cue only	−1.92	.059	
Spaced: cue–target vs. cue only	−5.49	<.001	0.30
Cue–target: massed vs. spaced	3.56	.001	0.19
Cue only: massed vs. spaced	−0.40	.687	

Note. Effect size (Cohen’s *d*) is reported only when the *t* value was significant

Appendix B

The complete results from ANOVAs of Experiment 2 are reported in Table B1. The results from follow-up simple effect tests of the interactions are reported in Table B2.

Table 8 Complete ANOVAs of Experiment 2

	Correct			Intrusion			No response		
	<i>F</i>	<i>p</i>	η_p^2	<i>F</i>	<i>p</i>	η_p^2	<i>F</i>	<i>p</i>	η_p^2
Proportion									
R	7.21	.009	.07	4.69	.033	.05	2.13	.148	
I	27.60	<.001	.23	7.24	.001	.07	11.41	<.001	.11
R × I	32.82	<.001	.26	12.86	<.001	.12	7.32	.001	.07
Latency									
R	79.45	<.001	.57	30.22	<.001	.31	1.01	.323	
I	6.98	.001	.11	4.92	.009	.07	4.34	.018	.13
R × I	2.30	.105		9.54	<.001	.12	2.84	.066	
I (massed)	1.26	.287							
I (spaced)	8.27	<.001	.11						
Proportion (Retrieval)									
R	360.34	<.001	.85						
I	1.95	.147							
R × I	<1								
Latency (Retrieval)									
R	132.88	<.001	.71						
I	2.37	.098							
R × I	1.15	.320							

Note. Effect size (η_p^2) is reported only when the *F* value was significant. For the significant *F* value, a Bonferroni test was used whenever necessary. R = repetition type (massed and spaced); I = item type (cue–target, cue only, and control)

Table 9 Simple effect tests following the interactions between repetition and item type of Experiment 2

	<i>t</i>	<i>F</i>	<i>p</i>	<i>d</i>	η_p^2
Proportion: Correct					
Massed: cue–target, cue only, & control		2.75	.066		
Spaced: cue–target, cue only, & control		50.80	<.001		.35
Cue–target: massed vs. spaced	−7.46		<.001	0.56	
Cue only: massed vs. spaced	2.05		.043	0.16	
Control: massed vs. spaced	−0.04		.967		
Proportion: Intrusion					
Massed: cue–target, cue only, & control		1.25	.290		
Spaced: cue–target, cue only, & control		16.78	<.001		.15
Cue–target: massed vs. spaced	5.04		<.001	0.33	
Cue only: massed vs. spaced	−0.58		.566		
Control: massed vs. spaced	−0.63		.533		
Proportion: No response					
Massed: cue–target, cue only, & control		1.93	.148		
Spaced: cue–target, cue only, & control		16.11	<.001		.15
Cue–target: massed vs. spaced	3.73		<.001	0.21	
Cue only: massed vs. spaced	−2.10		.039	0.12	
Control: massed vs. spaced	0.93		.354		
Latency: Intrusion					
Massed: cue–target, cue only, & control		<1			
Spaced: cue–target, cue only, & control		11.65	<.001		.14
Cue–target: massed vs. spaced	0.54		.593		
Cue only: massed vs. spaced	6.01		<.001	0.49	
Control: massed vs. spaced	4.58		<.001	0.47	

Note. Effect size (Cohen's *d* or η_p^2) is reported only when the *t* or *F* value was significant. For the significant *F* value, a Bonferroni test was used whenever necessary

Appendix C

Table 10 Complete results from statistical analyses of Experiment 3

	Correct			Intrusion			No response		
	<i>t</i>	<i>p</i>	<i>d</i>	<i>t</i>	<i>p</i>	<i>d</i>	<i>t</i>	<i>p</i>	<i>d</i>
Proportion	2.89	.004	0.43	−1.18	.240		−1.39	.167	
Latency	5.54	<.001	0.82	6.70	<.001	1.06	5.12	<.001	0.95
Proportion (Retrievable)	−17.65	<.001	2.63						
Latency (Retrievable)	5.92	<.001	0.90						

Note. Effect size (Cohen's *d*) is reported only when the *t* value was significant

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

References

- Atkinson, R. C., & Shiffrin, R. M., (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2, pp. 89–195). London, UK: Academic Press. [https://doi.org/10.1016/S0079-7421\(08\)60422-3](https://doi.org/10.1016/S0079-7421(08)60422-3)
- Bousfield, W. A., Sedgewick, C. H. W., & Cohen, B. H. (1954). Certain temporal characteristics of the recall of verbal associates. *American Journal of Psychology*, 67, 111–118. <https://doi.org/10.2307/1418075>
- Delaney, P. F., Verkoeijen, P. P. J. L., & Spiguel, A. (2010). Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 53, pp. 63–147). San Diego, CA: Elsevier Academic Press. [https://doi.org/10.1016/S0079-7421\(10\)53003-2](https://doi.org/10.1016/S0079-7421(10)53003-2)
- Diller, D. E., Nobel, P. A., & Shiffrin, R. M. (2001). An ARC-REM model for accuracy and response time in recognition and recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 414–435. <https://doi.org/10.1037/0278-7393.27.2.414>
- Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York, NY: Chapman & Hall.
- Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, 91, 1–67. <https://doi.org/10.1037/0033-295X.91.1.1>
- Heathcote, A., Popiel, S. J., & Mewhort, D. J. K. (1991). Analysis of response time distributions: An example using the Stroop task. *Psychological Bulletin*, 109, 340–347. <https://doi.org/10.1037/0033-2909.109.2.340>
- Hockley, W. E. (1984). Analysis of response time distributions in the study of cognitive processes. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 10, 598–615. <https://doi.org/10.1037/0278-7393.10.4.598>
- Hopper, W. J., & Huber, D. E. (2018a). Learning to recall: Examining recall latencies to test an intra-item learning theory of testing effects. *Journal of Memory and Language*, 102, 1–15. <https://doi.org/10.1016/j.jml.2018.04.005>
- Hopper, W. J., & Huber, D. E. (2018b). *Testing the PCR model of recall: Retrieval practice produces faster recall success but also faster recall failure*. Paper presented at the 59th annual meeting of the Psychonomic Society, New Orleans, LA.
- Huber, D. E., Tomlinson, T. D., Jang, Y., & Hopper, W. J. (2015). The search of associative memory and recovery interference (SAM-RI) memory model and its application to retrieval practice paradigms. In J. G. W. Raaijmakers, A. H. Criss, R. L. Goldstone, R. M. Nosofsky, & M. Steyvers (Eds.), *Cognitive modeling in perception and memory: A festschrift for Richard M. Shiffrin* (pp. 81–98). New York, NY: Psychology Press.
- Jang, Y., Lee, H., & Huber, D. E. (in press). How many dimensions underlie judgments of learning and recall redux: Consideration of recall latency reveals a previously hidden nonmonotonicity. *Journal of Mathematical Psychology*.
- Jang, Y., Wixted, J. T., & Huber, D. E. (2011). The diagnosticity of individual data for model selection: Comparing signal-detection models of recognition memory. *Psychonomic Bulletin & Review*, 18, 751–757. <https://doi.org/10.3758/s13423-011-0096-7>
- Jang, Y., Wixted, J. T., Pecher, D., Zeelenberg, R., & Huber, D. E. (2012). Decomposing the interaction between retention interval and study/test practice: The role of retrievability. *Quarterly Journal of Experimental Psychology*, 65, 962–975. <https://doi.org/10.1080/17470218.2011.638079>
- Kornell, N., Bjork, R. A., & Garcia, M. A. (2011). Why tests appear to prevent forgetting: A distribution-based bifurcation model. *Journal of Memory and Language*, 65, 85–97. <https://doi.org/10.1016/j.jml.2011.04.002>
- Kucera, H., & Francis, W. H. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University.
- Lacouture, Y., & Cousineau, D. (2008). How to use MATLAB to fit the ex-Gaussian and other probability functions to a distribution of response times. *Tutorials in Quantitative Methods for Psychology*, 4, 35–45. <https://doi.org/10.20982/tqmp.04.1.p035>
- Lewandowsky, S., & Farrell, S. (2010). *Computational modeling in cognition: Principles and practice*. Thousand Oaks, CA: SAGE. <https://doi.org/10.4135/9781483349428>
- MacLeod, C. M., & Nelson, T. O. (1984). Response latency and response accuracy as measures of memory. *Acta Psychologica*, 57, 215–235. [https://doi.org/10.1016/0001-6918\(84\)90032-5](https://doi.org/10.1016/0001-6918(84)90032-5)
- Malmberg, K. J., & Shiffrin, R. M. (2005). The “one-shot” hypothesis for context storage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 322–336. <https://doi.org/10.1037/0278-7393.31.2.322>
- Murnane, K., & Shiffrin, R. M. (1991). Word repetitions in sentence recognition. *Memory & Cognition*, 19, 119–130. <https://doi.org/10.3758/BF03197109>

- Nobel, P. A., & Shiffrin, R. M. (2001). Retrieval processes in recognition and cued recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 384–413. <https://doi.org/10.1037/0278-7393.27.2.384>
- Raaijmakers, J. G. W. (2003). Spacing and repetition effects in human memory: Application of the SAM model. *Cognitive Science*, 27, 431–452. [https://doi.org/10.1016/S0364-0213\(03\)00007-7](https://doi.org/10.1016/S0364-0213(03)00007-7)
- Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 14, pp. 207–262). New York, NY: Academic Press.
- Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, 88, 93–134. <https://doi.org/10.1037/0033-295X.88.2.93>
- Ratcliff, R., Clark, S., & Shiffrin, R. M. (1990). The list-strength effect: I. data and discussion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 163–178. <https://doi.org/10.1037/0278-7393.16.2.163>
- Ratcliff, R., & Murdock, B. B., Jr. (1976). Retrieval processes in recognition memory. *Psychological Review*, 83, 190–214. <https://doi.org/10.1037/0033-295X.83.3.190>
- Roediger, H. L., III, & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, 1, 181–210. <https://doi.org/10.1111/j.1745-6916.2006.00012.x>
- Rohrer, D. (1996). On the relative and absolute strength of a memory trace. *Memory & Cognition*, 24, 188–201. <https://doi.org/10.3758/BF03200880>
- Rohrer, D., & Wixted, J. T. (1994). An analysis of latency and interresponse time in free recall. *Memory & Cognition*, 22, 511–524. <https://doi.org/10.3758/BF03198390>
- Rotello, C. M., & Zeng, M. (2008). Analysis of RT distributions in the remember-know paradigm. *Psychonomic Bulletin & Review*, 15, 825–832. <https://doi.org/10.3758/PBR.15.4.825>
- Rouder, J. N., Lu, J., Speckman, P., Sun, D., & Jiang, Y. (2005). A hierarchical model for estimating response time distributions. *Psychonomic Bulletin & Review*, 12, 195–223. <https://doi.org/10.3758/BF03257252>
- Shiffrin, R. M. (1970). Memory search. In D. A. Norman (Ed.), *Models of human memory* (pp. 375–447). New York, NY: Academic Press.
- Shiffrin, R. M. (2003). Modeling memory and perception. *Cognitive Science*, 27, 341–378. [https://doi.org/10.1016/S0364-0213\(03\)00027-2](https://doi.org/10.1016/S0364-0213(03)00027-2)
- Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—Retrieving effectively from memory. *Psychonomic Bulletin & Review*, 4, 145–166. <https://doi.org/10.3758/BF03209391>
- Shiffrin, R. M., Ratcliff, R., & Clark, S. (1990). The list-strength effect: II. Theoretical mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 179–195. <https://doi.org/10.1037/0278-7393.16.2.179>
- Unsworth, N. (2015). The influence of encoding manipulations on the dynamics of free recall. *Memory & Cognition*, 43, 60–69. <https://doi.org/10.3758/s13421-014-0447-5>
- van den Broek, G. S. E., Segers, E., Takashima, A., & Verhoeven, L. (2014). Do testing effects change over time? Insights from immediate and delayed retrieval speed. *Memory*, 22, 803–812. <https://doi.org/10.1080/09658211.2013.831455>
- Wilson, J. H., & Criss, A. H. (2017). The list strength effect in cued recall. *Journal of Memory and Language*, 95, 78–88. <https://doi.org/10.1016/j.jml.2017.01.006>
- Wixted, J. T., Ghadisha, H., & Vera, R. (1997). Recall latency following pure- and mixed-strength lists: A direct test of the relative strength model of free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 523–538. <https://doi.org/10.1037/0278-7393.23.3.523>
- Zandbelt, B. (2014). exgauss: a MATLAB toolbox for fitting the ex-Gaussian distribution to response time data. *Computer software*. <https://doi.org/10.6084/m9.figshare.971318>