Journal of Experimental Psychology:
Learning, Memory, and Cognition
2005, Vol. 31, No. 2, 322–336

0278-7393/05/$12.00 DOI: 10.1037/0278-7393.31.2.322

The "One-Shot" Hypothesis for Context Storage

Kenneth J. Malmberg
Iowa State University

Richard M. Shiffrin
Indiana University Bloomington

In 3 experiments motivated by the implicit memory literature, the authors investigated the effects of different strengthening operations on the list strength effect (LSE) for explicit free recall, an effect posited by R. M. Shiffrin, R. Ratcliff, and S. E. Clark (1990) to be due to context cuing. According to the one-shot hypothesis, a fixed amount of context is stored when an item is studied for at least 1 or 2 s. Beyond the initial context storage, increases in study time or different orienting tasks do not influence the amount of context that is stored, and thus only spaced repetitions should produce a positive LSE. Consistent with prior findings, spaced repetitions always produced a positive LSE, but increases in depth of processing, study time, and massed repetitions did not. A model implements the one-shot hypothesis, and a role for context storage as a link between episodic and semantic memory is discussed.

In many theories of memory, the encoding "strength" is positively related to subsequent memory; the stronger an item is stored, the more likely that item will be remembered later. There are a number of strengthening operations. For example, increasing the amount of time that an item is studied (i.e., study time), increasing the number of times (i.e., repetitions) that an item is studied, and some orienting tasks improve explicit memory (Craik & Lockhart, 1972; Hintzman, 1974).

Mixed–pure list experiments determine what effect changing the strength of some memory traces has on the ability to remember items corresponding to other memory traces (e.g., Ratcliff, Clark, & Shiffrin, 1990). In these experiments, participants study lists of three kinds: all strong items (*pure strong*), all weak items (*pure weak*), and a mixture of some strong and some weak items (*mixed*). A *list strength effect* (LSE) is observed when the type of study list differentially affects memory for strong and weak items. Figure 1 shows that a *positive list-strength effect* obtains when weak items from mixed lists are more poorly remembered than weak items from pure lists, and strong items from mixed lists are better remembered than strong items from pure lists. Such a pattern is found if the strong list items on a mixed list inhibit, suppress, or otherwise interfere with memory of the weak items on the same list (and vice versa). A *negative list-strength effect* refers to the opposite pattern: mixed-weak items are better remembered than pure-weak items, and mixed-strong items are remembered worse than pure-strong items. That is, adding relatively strong traces to memory facilitates memory for relatively weak traces.

The direction of the LSE depends on the memory task (Ratcliff et al., 1990). A null or slightly negative LSE is usually found for recognition and paired-associates cued recall (Murnane & Shiffrin, 1991b; Ratcliff et al., 1990), but a positive LSE has always been observed for free recall (Ratcliff et al., 1990; Tulving & Hastie, 1972; Wixted, Ghadisha, & Vera, 1997). All of these studies strengthened items via spaced item repetitions, and investigators have not explored other ways to strengthen items (to the best of our knowledge). However, certain results from the implicit memory literature led us to suspect that a different result would obtain should strengthening be accomplished by massed study (i.e., increases in study time, the number of consecutive presentations, or depth of processing).[1] Specifically, we hypothesized that massed study operations would produce a null LSE for free recall, in contrast to the positive LSE always produced by spaced repetitions.

We present this reasoning and relevant results in the General Discussion. For present purposes, we simply note that we were led to suspect that context information might be stored only at the outset of a study trial and that the subsequent time spent studying would produce additional storage only of content information, such as meaning. If so, some theories predict different effects of massed and spaced study on the LSE in free recall. Thus, the purpose of these experiments is to investigate how different strengthening operations affect free recall in mixed–pure list experiments by varying levels of processing, study time, massed repetitions, and spaced repetitions.

The theories that motivate these studies are primarily the frameworks of the *search of associative memory theory* (SAM; Raaijmakers & Shiffrin, 1980, 1981; Shiffrin et al., 1990) and the *retrieving effectively from memory theory* (REM; Schooler, Shiffrin, & Raaijmakers, 2001; Shiffrin & Steyvers, 1997). Like most theories, they distinguish context and item information. There are

Kenneth J. Malmberg, Department of Psychology, Iowa State University; Richard M. Shiffrin, Department of Psychology, Indiana University Bloomington.

This research was supported in part by National Institute of Mental Health National Research Service Award Postdoctoral Fellowship 126431 to Kenneth J. Malmberg and National Institute of Mental Health Grant 12717 to Richard M. Shiffrin. We thank Amy Criss, Simon Dennis, and Michael S. Humphreys for many thoughtful discussions leading up to this research.

Correspondence concerning this article should be addressed to Kenneth J. Malmberg, Department of Psychology, Iowa State University, Ames, IA 50011. E-mail: malmberg@iastate.edu

---

[1] When convenient, we sometimes rather loosely refer to a period of time in which only a single item was studied as a massed repetition. Hence, an increase in *massed study* refers to an increase in study time or the number of consecutive presentations of an item. *Spaced repetitions* refers to two presentations of an item interpolated by the presentation of one or more different items.

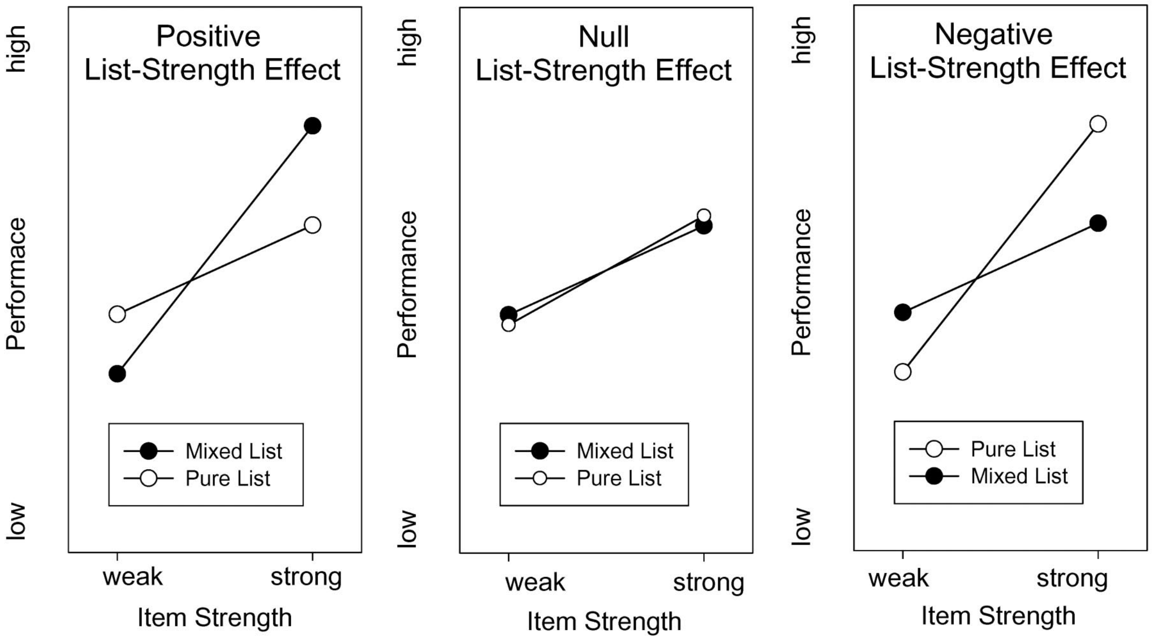

Figure 1. List strength effects.

various ways in which these types of information can be defined, and the two categories can have fuzzy boundaries (see *What Is Context?* in the General Discussion). For the present purposes, let *item* information represent meaning and other high-level codes particular to a given studied item; in typical tasks, such information lies within the focus of attention during the study episode. *Context* information is typically defined in terms of the physical, spatial–temporal, environmental, physiological, or emotional states in which the item was experienced (Murnane, Phelps, & Malmberg, 1999); in typical study situations, context information is probably not in the focus of attention. In most theories of episodic memory, probing memory with a mentally reinstated context cue allows for a task-relevant subset of the contents of long-term episodic memory to be accessed.

In SAM (the same account is used in REM), the positive LSE for free recall is posited to be a result of memory probes that use only context information (Shiffrin et al., 1990). Shiffrin et al. (1990) noted that free recall differs from recognition and cued recall in a critical respect (in the REM and SAM accounts): In free recall, some of the memory probes used during retrieval use context only and no item information (e.g., when initially attempting to recall or after an attempt to recall does not provide a more specific retrieval cue).

When context-only retrieval cues are used to probe memory, only one of the two activation factors comes into play: Images with strong context are activated more than those with weak context and hence are sampled preferentially. On such search cycles, a strongly positive LSE is predicted, because stronger nontarget images compete more effectively with the target. Other search cycles use context-plus-item cues, but this cue combination produces little or no LSE in any direction (Shiffrin et al., 1990; Shiffrin & Steyvers, 1997). Thus, the net effect of mixing the two kinds of search cycles is to produce a positive LSE overall. Shiffrin et al. (1990) made no distinction among the effects of different strengthening operations (e.g., increasing study time, depth of processing, or the number of massed or spaced repetitions) on the amount of context stored, and so the prediction of a positive LSE in free recall should hold for all.

However, there is at least one finding from the explicit memory literature that leads one to question the assumption that context storage increases roughly equally for all strengthening operations: It is found generally that testing items in an old context produces higher hit rates and false-alarm rates for single-item yes–no recognition than testing items in a new context (Murnane & Phelps, 1995; Murnane et al., 1999). This finding is predicted by global-matching theories of recognition because probing memory with an old context produces a greater level of familiarity than probing memory with new context, regardless of whether the item is a target or a foil (Murnane et al., 1999). In one experiment examining the effects of context change on recognition (Murnane & Phelps, 1995, Experiment 1), strengthening was accomplished by spacing repetitions in a given context; in a different experiment (Murnane & Phelps, 1995, Experiment 2), strengthening was accomplished by varying the amount of study time in a given context. In all other critical respects, the designs of the two experiments were the same. Murnane and Phelps found that strong items were better recognized than weak items in both experiments. However, the context effect increased as the number of spaced repetitions increased but not when study time increased.

This finding suggests that more context information is stored when items are strengthened by spaced repetitions than by massed study. One version of this explanation holds that each time an item is studied in massed fashion, a fixed amount, or "one shot," of context is stored in an episodic memory trace. A corollary to this view holds that spaced repetitions increase the strength of context storage because each sufficiently separated study episode produces storage of an additional shot of context. Although there are numerous ways that item and context information could be differentially stored as study conditions are manipulated, we consider only two simple classes of hypotheses:

> *Context-growth hypothesis:* Operations that increase the strength with which item information is stored on a given study trial also increase the storage of context information on that study trial.
>
> *One-shot context hypothesis:* There are conditions (such as lengthened study time) that increase the strength with which item information is stored but produce a fixed amount of context storage.

If repetitions of a given item are accumulated in a single trace, as REM and SAM assume (Shiffrin et al., 1990; Shiffrin & Steyvers, 1997), then according to both hypotheses, spaced repetitions at study will increase context storage in that trace. However, the context-growth and one-shot hypotheses make different LSE predictions if an item is studied only once, for varying durations. If the context-growth hypothesis is correct, and SAM or REM is the correct model, then all operations that increase item strength should also produce positive LSEs for free recall. If, however, the one-shot hypothesis is correct, then only spaced repetitions should produce a positive LSE for free recall. Our predictions concerning free-recall LSEs may be summarized as follows:

1. Study conditions that cause item strength and context strength both to rise will produce a positive LSE.
2. Study conditions that cause item strength but not context strength to rise will produce a null or slightly negative LSE.
3. Finally, note that according to REM and SAM, there should be a main effect of strength whenever either or both item and context strength are increased during study.

For those who are interested in the formal account, the REM and SAM models of free recall from which we derive these predictions are described in the Appendix.

Experiment 1: Levels of Processing

The type of activity performed when studying items has been shown in many studies to affect explicit memory (e.g., Craik & Tulving, 1975). An item is more likely to be remembered when studied in a manner that focuses on its meaning than when studied in a manner that focuses on its invariant physical characteristics (e.g., orthography, phonology). The former is often called "deep" or semantic-level processing and the latter, "shallow" or feature-level processing, and the manipulation is usually referred to as a "levels-of-processing" manipulation (Craik & Lockhart, 1972).

In this free-recall experiment, we use a levels-of-processing manipulation to create pure and mixed lists. All of the items on pure-weak and pure-strong lists are processed in a shallow or deep level, respectively, and on mixed lists, half of the items are studied in a shallow fashion and half in a deep fashion. The shallow orienting task involves detecting whether a word contains the letter *r*, and the deep orienting task involves making an animacy judgment ("alive" or "dead") to a studied word (cf. Nelson, 1977). Study is followed by instructions to freely recall as many list items as possible.

The predictions are those given at the end of the introduction: If deeper processing increases context storage, there should be a positive LSE. If deeper processing does not increase context storage, there should be a null or slightly negative LSE. In either case, items processed more deeply should be better recalled. A null or slightly negative LSE is consistent with the one-shot hypothesis, because this is what one expects if the deeper encoding task does strengthen context encoding to a greater degree than shallow encoding. The slight negativity might arise owing to probing with item information that is highly differentiated.

Method

Participants. One hundred one volunteers from introductory psychology courses at Indiana University participated in exchange for course credit.

Design and materials. Three types of lists were constructed: pure shallow, pure deep, and mixed. When shallow items were studied, participants were asked, "Is there an *r* in this word?" When deep items were studied, participants were asked, "Can this word be alive?" Manipulating the level of processing in this manner was chosen because it is known to significantly affect free-recall performance. On each study trial, the item was presented, and the participant was given 4 s to make a response.

The type of list studied was manipulated between participants to encourage the use of consistent strategies in accord with instructions. Thirty-five participants were randomly assigned to the pure-shallow condition, 36 to the pure-deep condition, and 30 to the mixed condition. In each condition, two 20-item lists composed of nouns selected from the Francis and Kučera (1982) norms were constructed for each participant. Half of the words on each list belonged to the class of animate objects, and half belonged to the class of inanimate objects; items were otherwise randomly assigned to conditions for each participant.

For all lists, study order was determined randomly. For mixed lists, depth of processing was blocked, and each participant studied one list in which the first half of the list was processed at a deep level and the second half at a shallow level and one list in which the blocks were reversed. Within each block, half of the items were animate objects with their order determined randomly. The presentation order for the two types of mixed lists was counterbalanced across participants.

Procedure. Participants were instructed that they would be presented a series of words to study, that they were to answer a question about each word, and that answering the question accurately was important, because the experiment was designed to investigate how answering different questions affects the ability to remember words. Participants in the pure-deep and pure-shallow conditions were told the nature of the question they were to answer. Participants in the mixed-list condition were told the nature of both types of questions and told that the type of question they would answer would switch halfway through each list. When the switch in questions occurred, a message indicating this fact was displayed for 10 s. For pure lists, this display indicated that the study list was half over.

On each trial, the to-be-remembered word appeared in the middle of a computer screen. Above the word was the question the participants were instructed to answer. Below the word were two circles, one labeled *alive* and one labeled *dead,* or one labeled *r* and one labeled *no r.* Responses were made with a mouse. Between study list and recall period, participants performed a simple arithmetic task for 30 s.

At test, participants were informed of the nature of memory tests, and they were instructed to attempt to recall in any order the words from the prior study list. Participants had 65 s to make their responses by typing them into the computer.

Results and Discussion

The findings are depicted in Figure 2. Participants who studied pure-deep lists recalled a greater proportion of words on average than those who studied pure-shallow lists (.31 and .20, respectively), $t(69) = 4.45$, $SEM = .025$. Participants who studied mixed lists also recalled a greater proportion of words processed deeply than words processed shallowly (.31 and .19, respectively), $t(29) = 4.11$, $SEM = .029$. Thus, the levels-of-processing manipulation served its purpose to differentially strengthen items. The interaction between list type (pure vs. weak) and level of processing was not significant ($F < 1.0$). Pure-shallow words were not recalled significantly better than mixed-shallow words (.20 and .19, respectively), $t(63) = 0.24$, and mixed-deep words were not recalled significantly better than pure-deep words (.31 and .31, respectively), $t(64) = 0.00$. Thus, a null LSE was observed.[2]

These results demonstrate that strengthening items via a levels-of-processing manipulation has a different effect on memory than strengthening items via spaced repetitions: Spaced repetitions produce a positive free-recall LSE, but levels-of-processing manipulations do not. The REM and SAM models predict such findings if spacing of repetitions increases the amount of context information stored in episodic images but increasing the depth of processing does not.

The key to the difference between spaced repetitions and levels of processing could well be the massing versus spacing of the study periods. If multiple presentations are necessary to increase the storage of context, then strengthening items by increasing study time should produce little or no LSE for free recall. We test this prediction in Experiment 2.

Experiment 2: Study Time Versus Spaced Repetitions

Murnane and Phelps (1995) found that context effects for recognition memory increase with increases in the number of spaced repetitions but not with increases in study time. This is explained by the one-shot hypothesis and is also consistent with results from Experiment 1. Therefore, the present experiment directly assesses the effect on the LSE for free recall of spaced repetitions versus study time. If spaced repetitions and increased study time both increase the storage of context information, a positive LSE should be seen in both cases. If spaced but not massed study increases context information, a positive LSE should be seen for the spaced condition only.

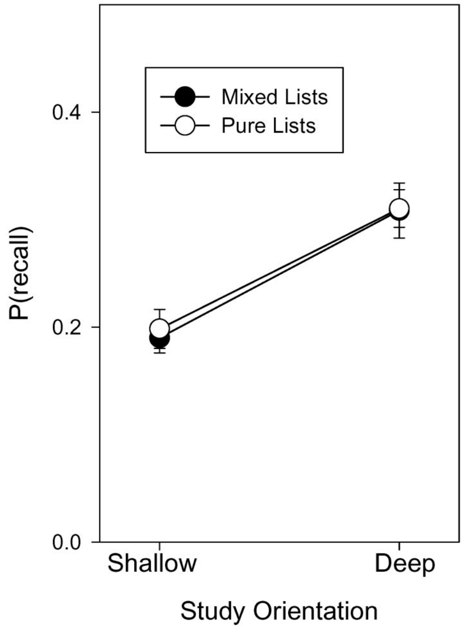

Figure 2. Experiment 1: probability of free recall for the four mixed–pure, shallow–deep list conditions. Error bars represent standard errors.

Method

Participants. Forty-four participants were paid $6 each and tested in individual booths on personal computers.

Design and materials. The design is illustrated in Figure 3. The design is patterned on those used by Murnane and Shiffrin (1991a, 1991b). List type (pure vs. mixed), item strength (strong vs. weak), and repetition type (massed vs. spaced) were manipulated as within-subject variables. Weak items were presented once for 1 s in both the massed and the spaced conditions. Strong items were presented once for 3 s in the massed condition and three times for 1 s in the spaced condition. The dependent measure was the proportion of words correctly recalled.

Pure lists consisted of either strong or weak items. Mixed lists were composed of equal numbers of strong and weak items. Two pure-weak lists, six mixed lists, and two pure-strong lists, each consisting of 20 different nouns with natural language frequencies of between 20 and 50 occurrences per million (Francis & Kučera, 1982), were constructed anew for each participant. The assignment of words to lists and conditions was determined randomly for each participant, as was list order.

Pure-weak lists were constructed and presented in the same manner for both the massed and the spaced-repetitions conditions. Items were studied for 1 s, and the order of presentation was determined randomly for each participant. Pure-weak lists do not vary in either study time or repetitions and are used as an indicator of baseline-level performance in both conditions.

In the massed condition, pure-strong lists were constructed and presented in the same manner as pure-weak lists except items were presented for 3 s. To prepare pure-strong lists in the spaced condition, we began by constructing two different 10-item pure-weak sublists. The first 30 serial positions were determined by concatenating three exact copies of one sublist, and the last 30 serial positions were determined in the same way using the remaining sublist. Therefore, each spaced item was presented three times at a lag of 10 items.

In mixed lists, strong and weak items were presented in separate blocks to minimize the possibility of rehearsal redistribution (cf. Murnane & Shiffrin, 1991b). The strong items preceded the weak items on two of the mixed lists and followed the weak items on the other mixed list.

In the massed condition, items in the first half or the second half of the 20-item study list were presented for 3 s, and items in the other half were presented for 1 s. The construction of a mixed list for the spaced condition proceeded in the same manner as the construction of a pure-strong list except only one of the 10-item sublists was repeated.

A distractor task, which involved adding a series of single digits, was used to minimize recency effects. The duration of the arithmetic task was varied so that the strong and weak item comparisons from pure and mixed lists were controlled for study–test delay (cf. Murnane & Shiffrin, 1991a).

---

[2] An alternative explanation is that participants in the mixed condition did not follow instructions and covertly made animacy judgments when they were supposed to do the letter identification task and that they covertly did the letter identification task when they were supposed to do the animacy judgment task. Such a scenario is possible, although we think it unlikely to have exactly offset the positive LSE that we would have otherwise observed. We thank Bill Batchelder for pointing this out to us.

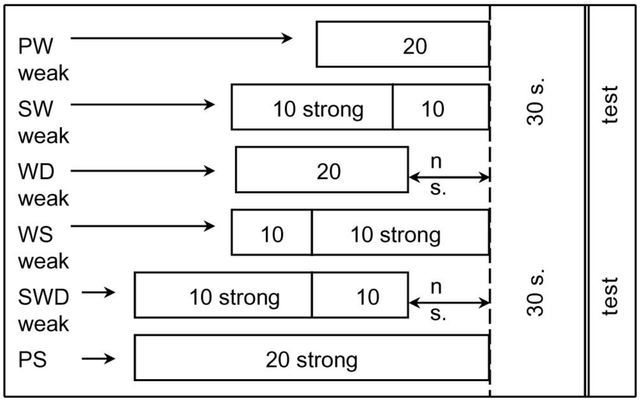

Figure 3. Design for Experiments 2 and 3. A row represents a list (PW = pure-weak list; SW = mixed list with weak items following strong items; WS = mixed list with strong items following weak items; PS = pure-strong list). The presentation begins at the onset of the leftmost bar in a row, and time runs toward the right-hand side until test (W represents weak). There was a 30-s arithmetic task between study and test for each list. Both the 30-s column and the rows labeled *n* were filled with an arithmetic task; in this case, *n* = 20 s. Thus, WD is a pure-weak delay list (i.e., a 50-s study–test interval), and SWD is a mixed delay list with weak items following strong items. The numbers refer to the number of each type of item on the list.

For instance, without controlling for study–test lag, the average lag is shorter for weak items on pure-weak than on mixed lists, and the average lag is longer for strong items on pure-strong than on weak lists. If the difference in study–test interval were to go unchecked, this could result in a trend toward a negative LSE independent of item strength. One pure-weak list, one strong–weak mixed list, the weak–strong mixed list, and the pure-strong list were followed by 30 s of arithmetic. An additional 20 s of arithmetic followed one pure-weak list (PW delay) and one strong–weak mixed list (SW delay) to control for study–test interval.

Procedure. Participants in both massed and spaced conditions were told that words would be presented to them one at a time and that their memory for the words would be tested after a brief math task. The nature of the free-recall task was also described. At test, participants were given 65 s to recall as many words from the study list as they could. Responses were entered into the computer using a keyboard. This cycle was repeated 10 times, once for each list, with the order of lists determined randomly for each participant.

Results and Discussion

The raw data are presented in Table 1. The first 10 items on the pure-weak list and the strong items from the strong–weak mixed list have no corresponding weak items matched in serial position, lag, and study–test delay on the mixed or the pure-strong lists, respectively (see Figure 3). Thus, the planned list-strength comparison for weak items was between the probability of recalling them from mixed lists versus the probability of recalling them from the PW delay list and the last 10 serial positions from the pure-weak list. These data were therefore combined to form the pure-weak observations for both massed and spaced conditions (i.e., pure-weak lists are equivalent for both massed and spaced repetitions). For the strong items, the planned comparisons were between the pure-strong items and the strong items from the SW delay and the weak–strong mixed lists. The strong-item data from these mixed lists were therefore combined to form the mixed-strong observations, done separately for the massed and spaced conditions.

Figure 4 shows that repetitions had a significant effect on the proportion of items correctly recalled. Items studied for 3 s were recalled more often than items studied for 1 s, $F(1, 43) = 124.3$, $MSE = .01$, and items studied three times spaced were recalled more often than items studied once, $F(1, 43) = 187.8$, $MSE = .02$.

If an LSE is observed, there should be an interaction between item strength and list type. Consistent with prior findings, a significant interaction was observed when items were strengthened via spaced repetitions, $F(1, 43) = 20.2$, $MSE = .01$. However, the interaction between item strength and list type was not reliable when items were strengthened via an increase in study time, $F(1, 43) = 1.19$, $MSE = .02$. Figure 4 shows that mixed-strong items were more likely to be recalled than pure-strong items when repetitions were spaced, $t(43) = 2.76$, $SEM = .03$, but not when they were massed, $t(43) = 0.37$, $SEM = .02$. Pure-weak items were more likely to be recalled than mixed-weak items when repetitions were spaced, $t(43) = 3.29$, $SEM = .01$, and when repetitions were massed, $t(43) = 3.05$, $SEM = .01$.

Thus, increases in study time produce only a partial LSE, whereas spaced repetitions produce a robustly positive LSE. These findings are approximately consistent with the findings from Experiment 1 and the one-shot hypothesis: One would not be far wrong in concluding from our theoretical analyses of the first two experiments that neither the amount of time a word is studied nor the task performed when studying a word significantly affects the amount of context information stored in that word's episodic trace. It seems that the amount of context stored is approximately determined by the number of times the item is presented. However, the partial LSE results of Experiment 2 suggest caution in drawing such a conclusion. It is possible that it takes more than 1 s to fully

Table 1
Experiment 2: Mean Percentage Correctly Recalled and the Standard Error of the Mean for Free Recall by List Type and Item Strength

	Repetition type			
	Study time		Spaced	
List and item type	P(C)	*SE*	P(C)	*SE*
Pure weak	.23	.01	NA	NA
Pure weak delayed	.23	.02	NA	NA
Mixed strong–weak				
Weak	.20	.02	.24	.02
Strong	.41	.03	.50	.03
Mixed strong–weak delayed				
Weak	.21	.02	.24	.02
Strong	.39	.03	.51	.03
Mixed weak–strong				
Weak	.20	.02	.17	.02
Strong	.35	.03	.52	.03
Pure strong	.36	.02	.44	.03

Note. P(C) = percentage correctly recalled. Pure lists consist of either strong or weak items, and mixed lists consist of half strong and half weak items. The delayed lists had a longer arithmetic task interpolated between study and test to control for study–test lag. NA indicates that pure weak lists vary in neither study time nor repetitions and are therefore used as an indicator of baseline performance for both conditions.

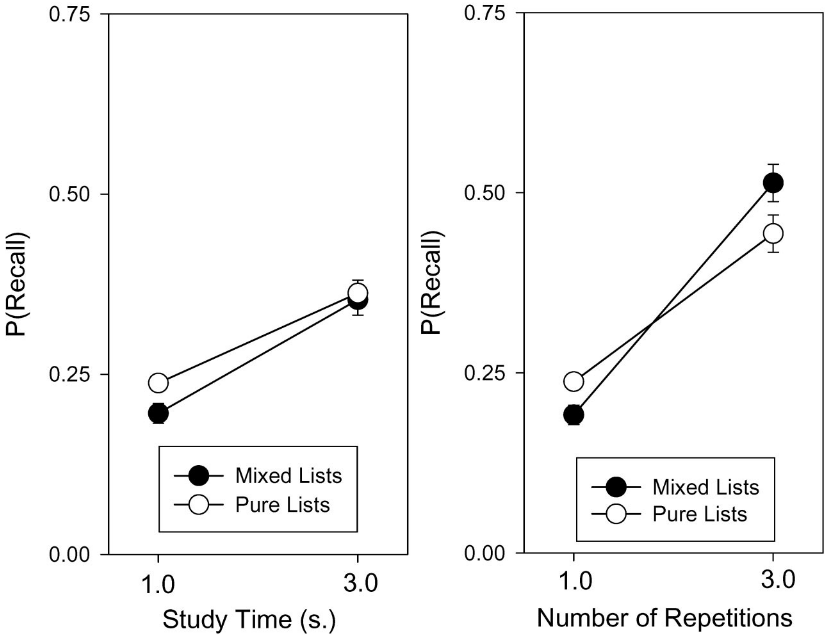

Figure 4. Experiment 2: probability of free recall for the four mixed–pure, strong–weak conditions, when items varied in study time (in seconds; left-hand side) or spaced repetitions (right-hand side). Error bars represent standard errors.

store the "one shot" of context, accounting for the partial LSE finding. This issue is explored further in Experiment 3.

Experiment 3: Study Time Versus Massed or Spaced Repetitions

This experiment tests a perhaps more realistic version of the one-shot hypothesis, one in which some minimum time is necessary to store a full "shot" of context. In particular, we test the hypothesis that it takes between 1 s and 2 s to complete the storage of context in an episodic trace. In Experiment 2, the study time for weak items was 1 s. This may not have been enough time to store a full shot of context in the weak traces, and thus a partially (maybe) positive LSE was observed.

This experiment has eight between-subjects conditions and two within-subject conditions. One between-subjects factor was the method for strengthening items: Items were strengthened via increases in study time, massed repetitions, spaced repetitions in the same order, or spaced repetitions in a different order. Each type of strengthening operation was used in two different between-subjects conditions (short vs. long study) in which the length of study was varied. Weak items were studied for 1 s in the short-study conditions and for 2 s in the long-study conditions. That is, when the strengthening operation was spaced repetitions, the duration of each presentation was either 1 s or 2 s in the short- and long-study conditions, respectively. When the strengthening operation was study time, weak items were studied for 1 s and 2 s and strong items were studied for 3 s and 6 s in the short- and long-study conditions, respectively.

By lengthening the minimum amount of study time, we increase the likelihood that a full shot of context is stored. Thus, we hypothesize that an LSE should be observed when repetitions are spaced, a small positive LSE should be observed in the study time condition when weak items are studied for 1 s (such as in Experiment 2), and no LSE should be observed in the study time condition when weak items are studied for 2 s.

We also varied whether spaced repetitions occurred in the same order or in a different order each time. The one-shot hypothesis under consideration here predicts a positive LSE regardless of the order in which spaced repetitions occur. Thus, the aforementioned predictions hold for both types of spaced repetitions.

Finally, items strengthened in massed study were strengthened by two different methods: massed repetitions and lengthened study time. For the massed repetitions condition, an item was presented three times in a row for either 1 s in the short-study condition or 2 s in the long-study condition. To the extent that massed and spaced repetitions have similar effects on memory, a positive LSE should be observed for massed repetitions. To the extent that massed repetitions and study time have similar effects on memory (which we anticipated), a small LSE should be observed for massed repetitions in the long-study condition, but a null LSE should be observed for massed repetitions in the short-study condition.

Method

Participants. Two hundred ninety-nine undergraduates enrolled in introductory psychology courses at Indiana University participated in individual booths equipped with personal computers in exchange for course credit. They were randomly assigned to the between-subjects conditions depending on the order in which they arrived at the lab.

Design and materials. The design followed that used in Experiment 2 and illustrated in Figure 3. Each of the eight miniexperiments manipulated list type (pure vs. mixed) and item strength (strong vs. weak) as within-subject variables. Participants studied six 20-item lists. One pure-weak list, one strong–weak mixed list, the weak–strong mixed list, and the pure-strong list were followed by 30 s of arithmetic. The recall tests were delayed by an additional 20 s or 40 s in the short and long conditions, respectively, following one pure-weak list (PW delay) and one strong–weak mixed list (SW delay).

The operation for strengthening items was varied between subjects. For the massed repetitions, an item was presented for study three times successively. For the spaced repetitions in the same order, items were repeated in the same order three times in spaced fashion (e.g., ABCABCABC). For the spaced repetitions in a different order, items were repeated in a different order three times in spaced fashion (e.g., ABCBEFGBH). The within-subject study time operation was the same used for Experiment 2. Each type of strengthening operation was used in two between-subjects conditions.

In addition, study time (short vs. long) was manipulated between subjects. In the between-subjects short-study condition, weak items were presented for 1 s, massed items were presented three times in a row for 1 s, strong items studied once were studied for 3 s, and spaced items were present three times for 1 s. In the between-subjects long-study condition, weak items were presented for 2 s, massed items were presented three times in a row for 2 s, strong items studied once were studied for 6 s, and spaced items were present three times for 2 s. The interstimulus interval was 150 ms.

The order in which spaced items were repeated was varied in the following manner: In the *same order* condition, spaced items were always studied before and after the same items (except for the first and last item presentations, e.g., A*B*C, A*B*C, A*B*C). In the *different order* condition, spaced items were always studied in a different order (e.g., A*B*C, D*B*E, F*B*G). Three different permutations of order were constructed to form each block of 10 items. The lag between repetitions in the same condition was 10, and an attempt was made to equate as nearly as possible the average lag in the different condition ($M = 9.1$). The specific orders used were [1,2,3,4,5,6,7,8,9,10], [3,8,2,5,1,7,10,6,9,4], and [3,1,10,2,9,5,8,6,4,7]. Thus, in the spaced–different condition the first item of a block was presented fifth and second on the second and third repetition cycles, respectively.

In all, 38, 37, 36, and 39 participants were assigned to the study time, massed, spaced–same, and spaced–different conditions, respectively, for the between-subjects short-study condition. A further 39, 40, 28, and 48 participants were assigned to the study time, massed, spaced–same, and spaced–different conditions, respectively, for the between-subjects long-study condition.

Procedure. The same procedure was used as in Experiment 2.

Results and Discussion

The raw data for the short- and long-study conditions are presented in Tables 2 and 3, respectively. These data were transformed in the same manner as in Experiment 2 (see above), and the results are illustrated in Figure 5, which plots the probability of recall as a function of list type and repetition type. The top row gives the data from the long-study condition (1 s vs. 3 s vs. 3 × 1 s), and the bottom row gives the data from the short-study condition (2 s vs. 6 s vs. 3 × 2 s).

For each of the eight between-subjects conditions, a repeated measures analysis of variance was performed, with item strength (weak vs. strong) and list type (mixed vs. pure) varied within subject. The results of these analyses are reported in Table 4. In each case, the main effect of the strengthening operation was significant: Strong items were better recalled than weak items (all ps $< .0005$). It is important to note that the interactions of item strength and list type were significant for all four spaced repetition conditions (all ps $< .0005$) and for the study time strengthening operation ($p < .025$). These interactions are indicative of LSEs, and we interpret these interactions in terms of their simple effects below. In no other condition did the interaction between item strength and list type approach traditional standards of reliability. These findings are inconsistent with the context-growth hypothesis, because increases in study time and massed repetitions did not produce positive LSEs; they are consistent with the one-shot hypothesis in that only spaced repetitions always produced reliable positive LSEs. They are also consistent with the hypothesis that it takes between 1 s and 2 s to store a full shot of context: A positive LSE was observed for the study time operation in the short-study

Table 2

Experiment 3: Short-Study Condition. Percentage Correctly Recalled and Standard Error of the Mean for Free Recall by List Type and Item Strength

	Repetition type							
	Study time		Massed		Spaced–same		Spaced–different	
List and item type	P(C)	*SE*	P(C)	*SE*	P(C)	*SE*	P(C)	*SE*
Pure weak	.22	.02	.22	.02	.21	.02	.22	.01
Pure weak delayed	.24	.03	.22	.02	.21	.02	.21	.02
Mixed strong–weak								
Weak	.21	.03	.24	.03	.18	.02	.20	.02
Strong	.44	.03	.41	.02	.45	.03	.46	.04
Mixed strong–weak delayed								
Weak	.23	.03	.23	.02	.23	.03	.20	.02
Strong	.38	.03	.41	.03	.48	.04	.39	.02
Mixed weak–strong								
Weak	.24	.03	.20	.03	.15	.03	.16	.02
Strong	.33	.04	.36	.03	.51	.04	.46	.03
Pure strong	.31	.03	.37	.02	.33	.02	.35	.02

Note. P(C) = percentage correctly recalled. Pure lists consist of either strong or weak items, and mixed lists consist of half strong and half weak items. The delayed lists had a longer arithmetic task interpolated between study and test to control for study–test lag. Spaced repetitions were made in either the same or a different order.

Table 3

Experiment 3: Long-Study Condition. Percentage Correctly Recalled and Standard Error of the Mean for Free Recall by List Type and Item Strength

	Repetition type							
	Study time		Massed		Spaced–same		Spaced–different	
List and item type	P(C)	*SE*	P(C)	*SE*	P(C)	*SE*	P(C)	*SE*
Pure weak	.29	.02	.28	.02	.24	.02	.24	.02
Pure weak delayed	.23	.02	.23	.02	.21	.02	.25	.02
Mixed strong–weak								
Weak	.28	.03	.19	.03	.27	.02	.23	.03
Strong	.46	.04	.50	.02	.55	.03	.49	.03
Mixed strong–weak delayed								
Weak	.25	.03	.28	.03	.19	.02	.24	.02
Strong	.38	.04	.44	.03	.51	.03	.47	.03
Mixed weak–strong								
Weak	.26	.03	.26	.03	.17	.02	.23	.03
Strong	.27	.03	.33	.03	.51	.03	.45	.03
Pure strong	.34	.03	.37	.02	.33	.02	.38	.02

Note. P(C) = percentage correctly recalled. Pure lists consist of either strong or weak items, and mixed lists consist of half strong and half weak items. The delayed lists had a longer arithmetic task interpolated between study and test to control for study–test lag. Spaced repetitions were made in either the same or a different order.

condition, but a null LSE was observed in the long-study condition.

We separate the analyses of the simple effects into the short- and long-study conditions. Consider first the short-study conditions (1 s vs. 3 s; top panels of Figure 5). Pure-weak items were more likely to be recalled than mixed-weak items in all cases, but the effect did not reach significance for the massed condition: For repetitions spaced in the same order, $t(35) = 3.18$, $SEM = .02$; for spacing in a different order, $t(38) = 2.92$, $SEM = .02$; for study time, $t(37) = 2.55$, $SEM = .01$; and for massed repetitions, $t(36) = 1.51$, $SEM = .02$, $p = .14$. Mixed-strong items were more likely to be recalled than pure-strong items in all cases, but the effect reached significance only for the spaced conditions: For repetitions spaced in the same order, $t(35) = 5.61$, $SEM = .03$; for spacing in a different order, $t(38) = 2.67$, $SEM = .03$; for study time, $t(37) = 1.30$; and for massed repetitions, $t(36) = 0.43$. Thus, there was a strong positive LSE for spaced repetitions, only a partial positive LSE for study time, and a null LSE for massed repetitions. These results replicate those from Experiment 2.

Let us turn now to the long-study condition (2 s vs. 6 s; bottom panels of Figure 5). Pure-weak items were more likely to be recalled than mixed-weak items only when repetitions were spaced: for same order, $t(38) = 2.23$, $SEM = .02$; for different order, $t(46) = 2.28$, $SEM = .02$; and massed, $t(40) = 2.17$, $SEM = .02$, $p < .04$, but not when study time was varied, $t(27) = -0.05$; in fact, the direction of the effect reversed. Mixed-strong items were more likely to be recalled than pure-strong items when repetitions were spaced: for same order, $t(38) = 4.39$, $SEM = .03$; for different order, $t(46) = 3.18$, $SEM = .02$, but not when massed, $t(40) = 0.29$, or when study time was varied, $t(27) = -0.69$; in fact, the direction of the effect again reversed. Thus, a strong positive LSE was observed when repetitions were spaced, massed repetitions produced a small (and possibly incomplete) positive LSE, and no LSE was observed for the study time condition.

If it takes between 1 s and 2 s to store a full shot of context, then the small LSE for items strengthened via an increase in study time that was observed in the short-study condition should have disappeared in this long-study condition, and this was the case. These results are consistent with the hypothesis that context storage is completed between 1 s and 2 s: For the massed repetitions condition, a small LSE was observed when weak items were studied for 1 s, but no LSE was observed when weak items were studied for 2 s. Thus, from these findings, we infer that the strength with which context is stored when items are studied for at least 2 s is approximately the same regardless of any additional amount of massed study, but that a small amount of context storage occurs between about 1 s and 2 s of study.

In no case did strengthening items via massed repetitions produce a reliable positive LSE. Thus, it appears that massed repetitions and increases in study time have similar effects on performance. There is a hint of a small effect for both the long- and short-study conditions, however. One reason for this may be that the onset of a new event or item may occasionally be sufficient to induce the storage of a small amount of context information. However, from these results it seems that this either occurs infrequently, produces little additional storage of context, or both.

General Discussion

In the Appendix, we describe a formal REM model of free recall that implements the assumption that a fixed amount of context is stored each time an item is studied for at least 2 s. The model's predictions based on the parameter values listed in the Appendix are presented in Figure 6. These predictions are a demonstration that the REM model that implements a version of the one-shot hypothesis is capable of handling our results.

What Is Context?

As with attention, everyone knows what context is: a theoretical construct necessary in order to describe how one remembers a single (usually recent) past occurrence of an item that may have

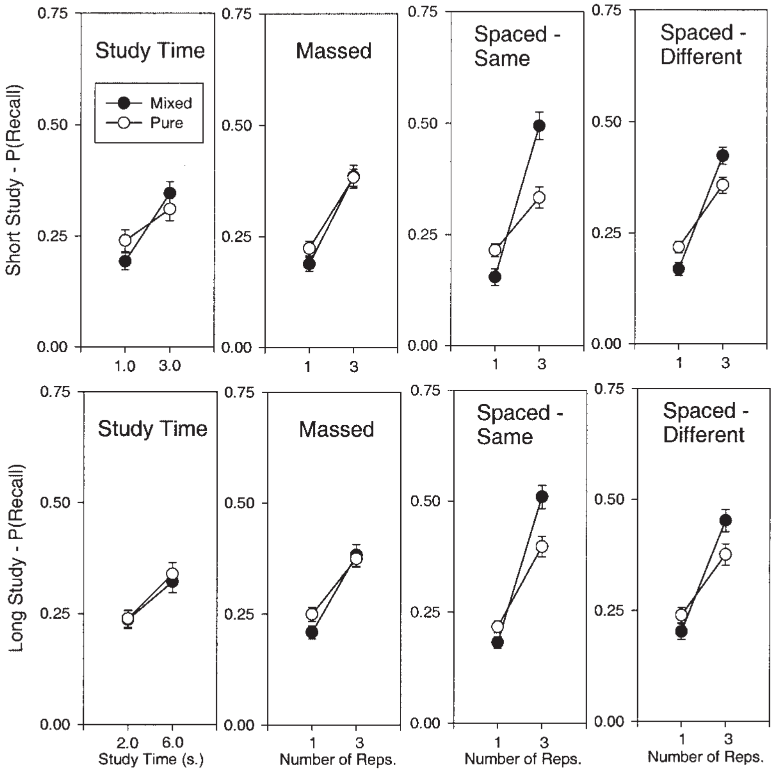

Figure 5. Experiment 3 data. Upper panels are for the shorter study time conditions; lower panels are for the longer study time conditions. Probability of free recall is depicted for the four mixed–pure, strong–weak conditions. The spaced presentation conditions are split into two conditions, one for when spaced repetitions always occurred in the same order and the other for when spaced repetitions occurred in a different order in each successive block of repeated words. s = seconds; Reps. = repetitions. Error bars represent standard errors.

been encountered thousands of times during life. Because context is not usually manipulated in an experiment, the construct often serves as a grab bag for all information other than the item information that is varied in the study. This state of affairs can lead to a very broad and fuzzy definition of context. Better and more specific definitions of context, including classification of the various kinds of context information into categories, require embedding the concept in a reasonably rich theoretical framework. At the least, such a framework requires distinguishing the representation of an item from the context in which it was encountered, even if that boundary might be blurry (cf. Howard & Kahana, 2002).

Context has been conceived in many ways. For instance, it has been conceived as a common node to which item representations are linked (Anderson & Bower, 1972), as a set of features that change randomly from study trial to study trial but more slowly than the features that represent items (Mensink & Raaijmakers, 1988; Shiffrin & Steyvers, 1997), as a set of features that change nonrandomly because they are generated by a set of oscillators of different frequencies (Brown, Preece, & Hulme, 2000), as a composite accumulation of the semantic properties of the studied items (Howard & Kahana, 2002), and as the environmental attributes that are present during encoding and as an integration of these environmental attributes with the abstract representation of the study item (Murnane et al., 1999). McGeoch (1942, pp. 501–505) anticipated many of these ideas:

> Everything learned is learned in response to stimulating or antecedent conditions which are a part of the learning situation and specific to it. It is learned also in a complex context of environing conditions not specific to it. There is the obvious external environment of stimuli impinging on the individual's exteroceptors, and there is the less obvious environment of intraorganic stimuli to the interoceptors. Correlated with these is the context of the individual's symbolic or ideational events. All are incidentally but inevitably present during practice, and it follows that the activities learned should be associated with some features of these environments. (p. 501)

However context is described, it is generally assumed by memory theorists to play a critical role in pointing at and limiting retrieval to a small, localized region of long-term memory. This assumption would be improved by differentiating the concept of context so as to specify the different kinds of context and the role each kind plays (see Murnane et al., 1999, for one example). Such developments are in their infancy but can be expected to accelerate.

The present experiments bear less on the issue of the kinds of context information than on the relation of context to item information. In particular, our results, when placed in the theoretical framework of SAM (Raaijmakers & Shiffrin, 1980) and REM models (Shiffrin & Steyvers, 1997), disconfirm the assumption that the strength with which item and context information is stored is highly positively correlated. At the least, our findings enlarge

Table 4
Analyses of Variance for Experiment 3

Condition	Effect	Repeated measures analysis of variance
Short study times (1 s vs. 3 s vs. 3 × 1 s)		
Study time	Item strength	$F(1, 37) = 38.8, MS^E = .012, p < .0005$
	pure vs. mixed	$F(1, 37) < .13, MS^E = .009, p = .72$
	interaction	$F(1, 37) = 5.48, MS^E = .011, p < .025$
Massed	Item strength	$F(1, 36) = 74.5, MS^E = .015, p < .0005$
	pure vs. mixed	$F(1, 36) = .70, MS^E = .006, p = .41$
	interaction	$F(1, 36) = 1.64, MS^E = .008, p = .21$
Spaced–same order	Item strength	$F(1, 35) = 148.7, MS^E = .013, p < .0005$
	pure vs. mixed	$F(1, 35) = 8.02, MS^E = .012, p < .01$
	interaction	$F(1, 35) = 43.2, MS^E = .010, p < .0005$
Spaced–different order	Item strength	$F(1, 38) = 173.8, MS^E = .009, p < .0005$
	pure vs. mixed	$F(1, 38) = .50, MS^E = .009, p = .48$
	interaction	$F(1, 38) = 14.2, MS^E = .009, p < .001$
Long study times (2 s vs. 6 s vs. 3 × 2 s)		
Study time	Item strength	$F(1, 27) = 16.7, MS^E = .014, p < .0005$
	pure vs. mixed	$F(1, 27) = .31, MS^E = .008, p = .59$
	interaction	$F(1, 27) = .23, MS^E = .009, p = .64$
Massed	Item strength	$F(1, 40) = 108.0, MS^E = .008, p < .0005$
	pure vs. mixed	$F(1, 40) = 1.23, MS^E = .008, p = .27$
	interaction	$F(1, 40) = 1.85, MS^E = .012, p = .18$
Spaced–same order	Item strength	$F(1, 35) = 178.8, MS^E = .014, p < .0005$
	pure vs. mixed	$F(1, 35) = 4.62, MS^E = .012, p < .05$
	interaction	$F(1, 35) = 28.8, MS^E = .007, p < .0005$
Spaced–different order	Item strength	$F(1, 46) = 112.9, MS^E = .009, p < .0005$
	pure vs. mixed	$F(1, 46) = 1.52, MS^E = .012, p = .23$
	interaction	$F(1, 46) = 20.1, MS^E = .008, p < .0005$

the importance of the theoretical distinction between item and context information by pointing at data that depend on this distinction.

According to the present REM model, item information and context information are distinguished in two ways (see Appendix for modeling details). First, context information is stored relatively early during the time course of study, and item information may be stored as long as the item remains in the focus of attention (as discussed below). Second, the context information stored in each episodic trace is highly similar (because the items are on the same list, and the list is the target of retrieval), whereas the item information contained in any two traces is usually highly dissimilar.

Note that our rather simple free-recall model predicts the same LSE regardless of whether spaced repetitions are in the same or in a different order, contrary to our data from Experiment 3 (which has recently been replicated within subjects in our lab).[3] We have too few relevant data to allow us to choose among explanations for this finding. Perhaps, the simplest explanation involves the difficulty of creating an effective context cue: Possibly a random list structure makes this more difficult than a repeated list structure. If so, a sampling advantage for strong items in a mixed list may be attenuated, producing a smaller LSE. Alternatively, participants may rely less on the use of a context-only cue when such cues are relatively ineffective and in turn rely more heavily on context-plus-item cues. It also is conceivable that the order of presentation affects the number and type of associations formed between items (although associations are not part of the present, simplified model). Last, some context features might not be independent of the items studied, and therefore repeating items in the same order strengthens existing contextual representations rather than storing more variable contextual representations (cf. Howard & Kahana, 2002). We decided not to augment the model in an attempt to fit the order effects, preferring to await the collection of additional data in future studies.

Why Is Only One Shot of Context Stored?

As a descendent of Atkinson and Shiffrin's (1968) modal model and SAM (Gillund & Shiffrin, 1984; Raaijmakers & Shiffrin, 1980, 1981; Shiffrin et al., 1990), REM assumes that memory performance is based on the interactions of a relatively permanent long-term memory store (LTS) and a limited capacity short-term memory store (STS). The STS temporarily maintains images so that they are readily available to control processes for the operations of cognition, including those control processes that encode images into LTS. The type and amount of information stored in LTS reflects the amount or type of controlled processing an item receives in the STS. In particular, when participants are trying to study for later memory tests, increased study, whether massed or spaced, increases content storage because attention is typically directed toward aspects of an item such as its meaning.

We hypothesize that context information is under the same circumstances stored incidentally, and perhaps in part automatically, because such information is not the object of attention. Thus, context information is stored once, at the time an item first enters STS, rather than as the result of a strategic decision. Later, encod-

---

[3] The data are available from Kenneth J. Malmberg on request.

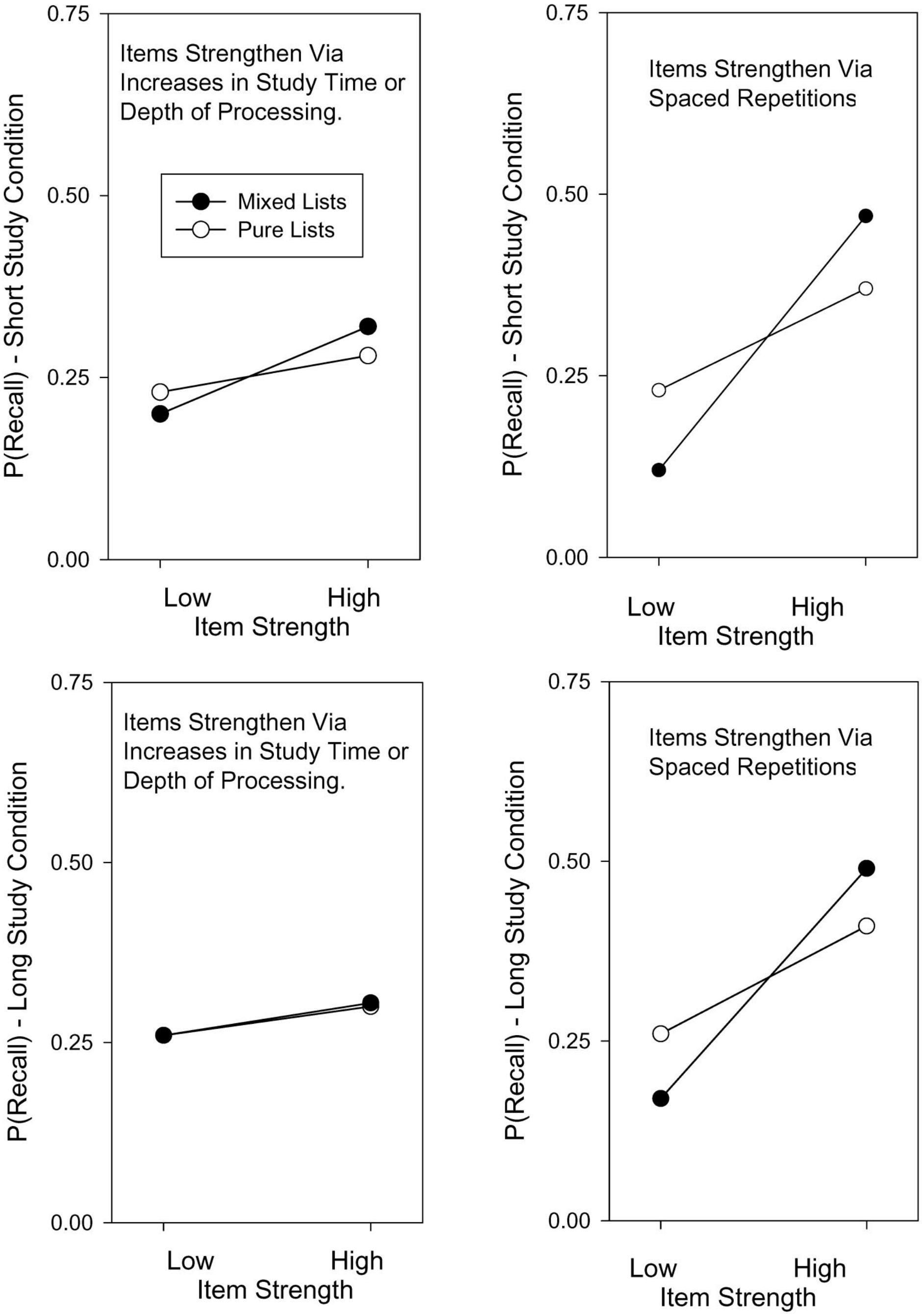

Figure 6. Qualitative predictions of retrieving effectively from memory theory.

ing turns to such things as meaning and associations. We further hypothesize that significant new storage of context occurs when an item has left STS and is then presented again, which occurs in the case of spaced repetitions. The use of the term *automatic* is not necessarily meant in the sense used in the attention literature. For example, incidental study might not necessarily produce as strong a shot of context as intentional study. Further, the degree of context storage might well change if different study tasks were used.

We note finally that a correlated reason why spaced presentation might produce additional context storage involves encoding variability and context drift (e.g., Anderson & Bower, 1972; Howard & Kahana, 2002; Mensink & Raaijmakers, 1988; Shiffrin & Steyvers, 1997). Although in the present model we fix context within lists, it is possible that context varies within lists, and if so, spaced repetitions would naturally increase context storage. Implicit in this hypothesis is the assumption that more variable context cues are more effective than relatively strong context cues, which is presumably the result of, say, storing one set of context features for 6 s versus 2 s. What the context variability hypothesis cannot explain is why massed study does not produce a robust positive LSE (or why the positive LSE was attenuated when repetitions occurred in a different order in Experiment 3). It is tempting to counter with the argument that massed study does not impart much context encoding after 2 s, which is what the one-shot hypothesis assumes, albeit in an extreme form. In summary, the one-shot hypothesis and context variability hypothesis are not necessarily mutually exclusive, but only the assumption that context is stored independently of item information and primarily in the first couple of seconds can account for the present findings.

In the context of these discussions it is useful to emphasize that the "one-shot" hypothesis is a qualitative approximation rather than a quantitative law: It seems likely to us that context storage for massed study rises sharply in the first second or so of study, and then falls off rapidly thereafter. Whether the context growth function continues to grow indefinitely (albeit at a slow rate), or grows to an asymptote, is a question the present studies do not address.

A Theoretical Link Between Implicit and Explicit Memory

The present studies, and the one-shot hypothesis, were originally motivated by a set of findings indicating that some implicit benefits do not increase when items are studied for longer periods of time. This result is one of several empirical dissociations between "perceptually driven" implicit and explicit memory. In particular, repetition effects are found for many tasks, including free recall, paired-associate cued recall, and recognition (see Hintzman, 1974, for a review). The massed repetition effect is closely related to another robust phenomenon, that extra study time improves memory. In addition, level-of-processing operations also improve recall and recognition (e.g., Craik & Tulving, 1975). The typical memory tasks for which these generalizations hold require the participant to remember a specific contextually defined event and have become known as direct or explicit memory tasks (Richardson-Klavehn & Bjork, 1988).

Another class of tasks may be accomplished by using general knowledge (e.g., perceptual identification and word-fragment completion) and do not require memory for any single past event but nonetheless reveal the occurrence of recent events because those events alter task performance. These have become known as indirect or implicit tasks (Richardson-Klavehn & Bjork, 1988). Most commonly, a prior experience with an item improves the subsequent performance of indirect memory tests, even though the indirect task may be accomplished without reference to the first presentation, and even though the participant may be unaware that the tested item has been presented in a first phase (e.g., Richardson-Klavehn & Bjork, 1988). For example, identification accuracy (to a brief exposure) and identification speed (to a clear exposure) are improved following prior exposure to a word.

In contrast to explicit results, massed and spaced repetitions produce different effects on perceptually driven implicit memory such as perceptual identification and naming: Implicit memory usually improves following spaced repetitions (e.g., Greene, 1990; Perruchet, 1989) but usually does not improve beyond the initial presentation for massed repetitions (e.g., Challis & Sidhu, 1993; Roediger & Challis, 1992). Similarly, extra study time (Wolters & Prinsen, 1997) and levels-of-processing instructions (e.g., Graf, Mandler, & Haden, 1982) typically do not improve indirect memory performance. The key finding to be explained, therefore, is the fact that extra massed study, or extended study, produces better direct or explicit memory but does not improve the performance of many perceptually indirect memory tasks.

The REM account of such dissociations runs as follows (e.g., Schooler et al., 2001; Shiffrin & Steyvers, 1997). When a word is studied, some information is added to the studied word's lexical–semantic image, as long as that information is not already stored in the trace. Thus, some relatively novel current context information, and some novel low-level physical information such as font, screen color, and so on, is added to the word's lexical–semantic trace, but semantic information is not added. During a perceptually driven implicit memory test, it is assumed that at least some current context information, including low-level feature information (e.g., font), joins the content information in the memory probe (even though the task may not require such information to be accomplished successfully). Extra matching is then produced between probe and lexical–semantic trace by the matching of the context in the probe and the low-level physical information in the probe to the same features that had been stored during study. This extra matching is what, according to REM, produces the implicit benefit (see also Jacoby & Dallas, 1981).

In this account, adding more semantic features to episodic memory does not affect priming, because those semantic features are not part of the probe in a perceptual implicit memory task. Adding more context features at study will produce more priming (the extra context features might be added to both episodic and implicit traces, although only the addition to the lexical–semantic traces produces priming). However, the one-shot hypothesis, along with the present results from explicit memory, implies that only spaced study produces such increased context storage; hence, extra study should produce greater implicit benefit in perceptual implicit memory tasks only for spaced and not massed study. Note finally that a close connection between the implicit and explicit accounts would require that storage conditions that do lead to increased explicit context storage also produce increased priming. In particular, our present results suggest that the use of brief study times (under a second, say) might produce increased priming for increased massed study. Rajaram, Srinivas, and Travers (2001) reported a result that might be relevant: Very brief study (in a Stroop task) was combined with a manipulation of attention to the content or color of the studied words (later implicit tests always used gray stimuli). Attention to content produced more priming, which we suggest could be due to additional context storage.

According to the above account of implicit benefits, any experimental manipulation that increases storage of context information in lexical–semantic images should increase the implicit benefit. The fact that extra study time, massed repetitions, and levels-of-processing instructions do not increase the implicit benefit therefore implies that these manipulations do not increase the storage of context information in lexical–semantic images. It was this reasoning that led us to hypothesize that these same manipulations might not increase the storage of context information in explicit images and led us to carry out the present set of experiments.

Conclusions

In these experiments, we examined strength and list strength effects in free recall. Increases in spaced repetitions, massed repetitions, study time, and depth of processing produced better recall for the items strengthened, but spaced repetitions produced a qualitatively different LSE than the other manipulations: Spaced repetitions produce a robust positive LSE, but increases in study time, massed repetitions, and depth of processing produce little or no LSE. That is, strong items tend to suppress or inhibit recall of weaker items only when strengthened via spaced repetitions.

It may be of interest to some readers that our theoretical account of the complex findings in the present article was not generated after the fact. Several implicit memory findings and Schooler et al.'s (2001) implicit memory model led us to predict in advance that the present set of interactions would occur by adapting within the REM framework the SAM model for free recall (e.g., Raaijmakers & Shiffrin, 1980, 1981; Shiffrin & Steyvers, 1998), the Shiffrin et al. (1990) account of LSEs, and the additional assumption that we term the *one-shot hypothesis*. In sum, the one-shot hypothesis provides an explanation for why explicit and implicit memories are sometimes differently and sometimes similarly affected by a variety of strengthening operations.

References

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review, 79,* 97–123.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In W. E. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2, pp. 89–195). New York: Academic Press.

Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review, 107,* 127–181.

Challis, B. H., & Sidhu, R. (1993). Dissociative effect of massed repetition on implicit and explicit measures of memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 115–127.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11,* 671–684.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and retention of words in episodic memory. *Journal of Experimental Psychology: General, 104,* 268–294.

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.

Francis, W. N., & Kučera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar.* Boston: Houghton Mifflin.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91,* 1–67.

Graf, P., Mandler, G., & Haden, P. (1982). Simulating amnesic symptoms in normal subjects. *Science, 218,* 1243–1244.

Greene, R. L. (1990). Spacing effects on implicit memory tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 1004–1011.

Hintzman, D. L. (1974). Theoretical implications of the spacing effect. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola Symposium* (pp. 77–99). Hillsdale, NJ: Erlbaum.

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology, 46,* 269–299.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 110,* 306–340.

McGeoch, J. A. (1942). *The psychology of human learning.* New York: Longmans, Green, and Co.

Mensink, G. J., & Raaijmakers, J. G. W. (1988). A model of interference and forgetting. *Psychological Review, 95,* 434–455.

Murnane, K., & Phelps, M. P. (1995). Effects of changes in relative cue strength on context-dependent recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 158–172.

Murnane, K., Phelps, M. P., & Malmberg, K. (1999). Context-dependent recognition memory: The ICE theory. *Journal of Experimental Psychology: General, 128,* 403–415.

Murnane, K., & Shiffrin, R. M. (1991a). Interference and the representation of events in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 855–874.

Murnane, K., & Shiffrin, R. M. (1991b). Word repetitions in sentence recognition. *Memory & Cognition, 19,* 119–130.

Nelson, T. O. (1977). Repetitions and depth of processing. *Journal of Verbal Learning and Verbal Behavior, 16,* 151–172.

Perruchet, P. (1989). The effect of spaced practice on explicit and implicit memory. *British Journal of Psychology, 80,* 113–130.

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: Advances in research and theory (Vol. 14, pp. 207–262). New York: Academic Press.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88,* 93–134.

Rajaram, S., Srinivas, K., & Travers, S. (2001). The effects of attention on perceptual implicit memory. *Memory & Cognition, 29,* 920–930.

Ratcliff, R., Clark, S. E., & Shiffrin, R. M. (1990). List-strength effect: I. Data and discussion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 163–178.

Richardson-Klavehn, A., & Bjork, R. A. (1988). Measures of memory. *Annual Review of Psychology, 39,* 475–543.

Roediger, H. L., & Challis, B. H. (1992). Effects of exact repetition and conceptual repetition on free recall and primed word-fragment completion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 3–14.

Schooler, L., Shiffrin, R. M., & Raaijmakers, J. G. W. (2001). A Bayesian model for implicit effect in perceptual identification. *Psychological Review, 108,* 257–272.

Shiffrin, R. M., Ratcliff, R., & Clark, S. E. (1990). List-strength effect: II. Theoretical mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 179–195.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin & Review, 4,* 145–166.

Shiffrin, R. M., & Steyvers, M. (1998). The effectiveness of retrieval from memory. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 73–95). London: Oxford University Press.

Tulving, E., & Hastie, R. (1972). Inhibition effects in intralist repetitions in free recall. *Journal of Experimental Psychology, 92,* 297–304.

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.

Wolters, G., & Prinsen, A. (1997). Full versus divided attention and implicit memory performance. *Memory & Cognition, 25,* 764–771.

Appendix

Retrieving Effectively From Memory (REM) and the One-Shot Hypothesis

We describe a computer simulation of a relatively simple version of REM that implements the one-shot hypothesis and from which we can generate concrete predictions. This model is applied to Experiment 3 only, although the description makes obvious what would be the treatment of the other studies.

Lexical–Semantic Images

In REM, a word in lexical–semantic memory is represented as a large vector of feature values, including both context and content features. The value **V** determines the mean and variability of features in the environment, according to a geometric distribution:

$$P[\mathbf{V} = j] = (1 - g)^{j-1}g, \text{ where } j = 1, \ldots, \infty. \quad \text{(A1)}$$

The geometric distribution is an arbitrary choice made for the sake of simplicity. The important thing to note is that REM assumes that some feature values are more common in the environment than others. As is discussed further below, relatively rare features (greater feature values) provide more diagnostic matching information than relatively common feature values.

Encoding of Episodic Images

Episodic images represent particular events and tend to be stored incompletely and inaccurately. When an item is repeated in a similar context, the new information is typically accumulated in the previous episodic image, so that repetitions tend to be represented in one trace, but a separate trace can also be stored, especially when the contexts differ significantly (e.g., Murnane & Shiffrin, 1991b). In the present context, accumulation means switching zero values (nothing stored) with nonzero feature values. REM assumes that nonzero feature values are never switched (e.g., corrected) without the application of attention or effort to do so. Episodic images contain information about items and information about the context in which the items were experienced. At each study of an item, REM assumes a tendency for storage in episodic images of both item information and current context information (Mensink & Raaijmakers, 1988; Shiffrin & Steyvers, 1997).

We assume that when a word is studied, its lexical–semantic image is accessed and the content information found in that vector is entered into a short-term rehearsal buffer (see, e.g., Atkinson & Shiffrin, 1968) where it joins features representing the physical characteristics of that situation and other subjective context information. Suppose the lexical–semantic vector representing an item contains w_{I} content features and w_{c} context features. Assume that a context vector also of length w_{c} exists for each list of items studied. When two or more lists of items are studied, the context vectors for adjacent lists ought to be similar, because context is assumed to change between lists, but gradually and continuously (Mensink & Raaijmakers, 1988). This can be achieved by randomly choosing context features to be changed for a given point in time; the more features are selected, the greater is the rate of context change. However, the present model is restricted for simplicity to images from the most recent list (see Diller, Nobel, & Shiffrin, 2001, for a treatment of the more general case).

Within lists, all repetitions of an item are assumed for simplicity in the present analyses to produce a single episodic image, regardless of the kind of spacing. We assume this single episodic image is composed of w_{I} feature values representing a partial and inaccurate copy of the item features in that item's lexical–semantic vector and w_{c} feature values representing a partial and inaccurate copy of the current context.

The probability of storing a feature in an episodic image increases as the number of attempts to store it increase. For each of t storage attempts, there is a probability u^* of storing a nonzero feature value (correct or incorrect); if nothing is stored during these attempts, then a zero is stored (representing no information about that feature). After t storage attempts, therefore, the probability that a nonzero feature will be stored is $1 - (1 - u^*)^t$. If a nonzero feature value is stored, there is a probability c that it is stored correctly, and otherwise the stored feature value is chosen randomly from the base-rate distribution (Equation A1), although it is possible that the "correct" value will be chosen when a random draw is made from the appropriate geometric distribution. These rules produce an episodic image that is an incomplete and error-prone copy of the lexical–semantic and context vectors held in the rehearsal buffer. In this model, we use the REM assumption that a feature (context or content) that is already stored is not altered by future storage attempts (whether massed or spaced), even if the original storage was incorrect (Shiffrin & Steyvers, 1997).[A1]

We now describe the rules that govern the storage of *content* (or "item") features. The amount of time an item is studied and the nature of spacing jointly determine the number of attempts made to store content features. Consider massed study first: The number of attempts, t_j, at storing a content feature for an item residing continuously in the rehearsal buffer for j units of time is computed from the following equation:

$$t_j = t_{j-1}(1 + e^{-aj}), \quad \text{(A2)}$$

where a is a rate parameter and t_1 is the number of attempts at storing a feature in the first 1 s of study. Here, we assume that $j > 1$ corresponds to study time in seconds, but this could obviously be implemented at different levels of precision. Massed repetitions are treated identically to a single presentation of the same total duration. Thus, increased study time or massed repetitions increase the storage of content information, but the gain in information diminishes as the item is studied longer. We assume additionally that "deep" processing increases the storage of content information: The number of storage attempts for shallow processing is t_{s} and for deep processing is t_{d}, where $t_{\mathrm{s}} < t_{\mathrm{d}}$.

Spaced repetitions are assumed to produce a single episodic image rather than multiple images. If an item has left rehearsal and is presented again, additional attempts are made to store content features in that single image, but at a higher rate than would have been predicted on the basis of the diminishing returns of Equation A2. In particular, each spaced repetition of duration j seconds allows for t_j additional attempts to store content features (where t_j is determined according to Equation A2).

Now consider the storage of *context* features. The one-shot hypothesis is implemented by assuming the following:

1. Extra massed repetitions, extra study time, and relatively deep processing do *not increase* substantially the storage of context information in episodic images. More precisely, any continuous period of time greater than 2 s that an item is in the rehearsal buffer results in t_{C} attempts to store context features, regardless of total time or depth of processing.
2. Spaced repetitions *increase* the storage of context information in episodic images. Each spaced repetition of an item of at least 2 s duration produces t_{C} additional storage attempts for context features.

[A1] Shiffrin and Steyvers (1997) assumed that previously stored features could be altered, but only through the application of "attention" to the features in question. This assumption allowed correction over developmental time of incorrectly stored features in lexical–semantic features.

(*Appendix continues*)

The SAM–REM Model for Free Recall

The model we use for free recall is a particularly stripped down version of that normally applied to free recall (e.g., Raaijmakers & Shiffrin, 1980, 1981; Shiffrin & Steyvers, 1998). We assume only the images of items on the recent list are involved in the retrieval operations and that all such images are involved. We assume that all probes of memory use only the context cue. The present context is matched to the episodic images of items in the most recent list. Each such image is assigned a likelihood ratio on the basis of the context features that match, and their values, and those that do not match. The equation for calculating the likelihood ratios is as follows:

$$\lambda_j = (1 - c)^{n_{ij}} \prod_{i=1}^{\infty} \left[\frac{c + (1 - c)g(1 - g)^{i-1}}{g(1 - g)^{i-1}} \right]^{n_{ijm}}, \qquad \text{(A3)}$$

where g is the long-run environmental base rate for the occurrence of features (i.e., g[system]), i is a context feature value ranging from 1 to infinity, n_{ij} is the number of mismatching context features in I_j, and n_{ijm} is the number of times context feature i matched the retrieval cue with value j.

Free recall operates as a memory search, with cycles of sampling and recovery. The probability of sampling image, I_i, given the context retrieval cue, Q, is as follows:

$$P(I_i|Q) = \frac{\lambda_i^{\gamma}}{\sum \lambda_k^{\gamma}}, \qquad \text{(A4)}$$

where λ_i is the likelihood ratio for image I_i from Equation A3 and γ is a scaling parameter (needed because the likelihood ratios are highly skewed; see Shiffrin & Steyvers, 1998). The denominator is the sum of the scaled likelihood ratios across the activated images. The more similar the context features in I_i are to Q, the more likely it will be sampled.

When repetitions are spaced, context strengths are stronger for the repeated than for the nonrepeated items, so the likelihood ratios in mixed lists will tend to be a mixture of large and small values. There will therefore be a tendency to sample the strong images at the cost of sampling the weaker images, in comparison with pure lists, in which sampling of all images is on average equally likely. This sampling difference is the cause of the observed positive LSE in free recall. Conversely, when repetitions are massed, context strengths are equal for all items, so the difference between pure and mixed lists disappears, and the LSE becomes null.

In this simplified model, as in more complete versions of SAM and REM, item features play a role once an image is sampled. We assume that the probability that a sampled image is recovered and output, P(R), is a monotonically increasing function of the proportion of correctly stored item-content features in the sampled image:

$$P(\mathrm{R}) = \rho_r^{\tau}, \qquad \text{(A5)}$$

where ρ_r represents the proportion of correctly stored item features in that image and τ is a parameter (for more realistic but more complicated recovery functions, see Diller et al., 2001). Because repeated items have more completely encoded item features than weak items, regardless of type of spacing, repeated items are more likely to be recovered once sampled, producing a main effect of strength: Stronger items are recalled better regardless of spacing manipulations. Retrieval continues until $K_{\max}$ failures to output an item occur.

At this point, we should mention that the free recall model deriving from Raaijmakers and Shiffrin (1980, 1981) and implemented in REM (as discussed in Shiffrin & Steyvers, 1998) is quite a bit more complicated than the version presented here. In particular, some cycles of the search use context probes only, and others use context-plus-item probes. When these two cues are combined on a given search cycle, differentiation lowers the likelihood ratio for images with stronger item information and increases the likelihood ratio for images with stronger context information. Thus, for spaced repetitions, these two factors tend to trade off, and an approximately null LSE is predicted (see Shiffrin et al., 1990, for a full discussion). A more complete version of the free-recall model applied to the case of spaced repetitions therefore would involve a mixture of search cycles on which a null LSE occurs (context-plus-item cuing) and other cycles on which a positive LSE occurs (context cuing only). Such a mixture produces a positive LSE in net. Conversely, for massed study, context does not increase, and so joint probes with context and item cues will be governed by the differentiation factor alone, producing a slightly negative LSE for such probes. In principle, this means that for massed study in the more complex model, a very small amount of increased context storage could have occurred, with slightly positive LSEs on the context probes canceling slightly negative LSEs on the joint probes.

Parameter Values for the Free-Recall Simulation

Following are the parameters and the basis for the choice of values (most were carried over from Shiffrin & Steyvers, 1997, and are denoted *S&S;* a few were chosen after some brief exploration of alternative values). Predictions when items are strengthened via a levels-of-processing and massed repetitions manipulation were generated from the same parameter values as those used to generate predictions for increasing study time.

g(system): The value used by the system to calculate likelihood ratios (from S&S); $\hat{g}_s = .40$.

g (Experiment 1 high-frequency words): The value used to construct the lexical entries for content features (from S&S); $\hat{g} = .45$.

w_{I}: Number of content features in each lexical vector (from S&S); [circ]$w_{\mathrm{I}} = 20$.

w_{c}: Number of context features in each lexical vector (from S&S); [circ]$w_{\mathrm{c}} = 20$.

c: The probability of correct storage of a feature value given something is stored (from S&S); [circ]$c = .80$.

u^*: The probability of succeeding at storing something at each of t storage attempts (estimated after a few values were tried); $\hat{u}^* = .02$.

t_1: Number of storage attempts in the first 1 s of study (estimated after a few values were tried); $t_1 = 6$.

a: The scaling parameter in Equation A2 governing storage attempts over time (simply set to 1.0); $a = 1.0$.

τ: The scaling parameter for recovery (set after a few values were tried); $\tau = 0.5$.

γ: The scaling parameter for likelihood ratios during sampling (set after a few values were tried); $\gamma = 0.2$.

$K_{\max}$: The number of failed cycles determining end of the search (set after a few values were tried); $K_{\max} = 16$.

Received February 15, 2002
Revision received July 19, 2004
Accepted July 20, 2004 ■