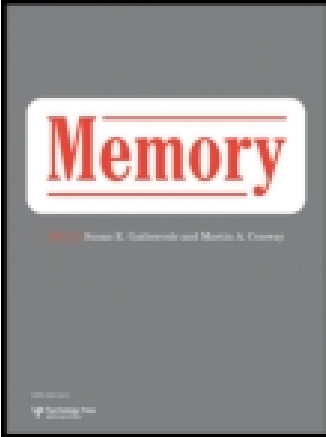


This article was downloaded by: [Duke University Libraries]

On: 21 May 2015, At: 09:26

Publisher: Routledge

Informa Ltd Registered in England and Wales Registered Number:
1072954 Registered office: Mortimer House, 37-41 Mortimer
Street, London W1T 3JH, UK



Memory

Publication details, including
instructions for authors and subscription
information:

[http://www.tandfonline.com/loi/
pmem20](http://www.tandfonline.com/loi/pmem20)

The Spacing Effect Depends on an Encoding Deficit, Retrieval, and Time in Working Memory: Evidence

Kelly Braun David C. Rubin

Published online: 15 Oct 2010.

To cite this article: Kelly Braun David C. Rubin (1998) The Spacing Effect
Depends on an Encoding Deficit, Retrieval, and Time in Working Memory:
Evidence, *Memory*, 6:1, 37-66, DOI: [10.1080/741941599](https://doi.org/10.1080/741941599)

To link to this article: <http://dx.doi.org/10.1080/741941599>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of
all the information (the "Content") contained in the publications
on our platform. However, Taylor & Francis, our agents, and our
licensors make no representations or warranties whatsoever as
to the accuracy, completeness, or suitability for any purpose of

the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

The Spacing Effect Depends on an Encoding Deficit, Retrieval, and Time in Working Memory: Evidence from Once-presented Words

Kelly Braun and David C. Rubin

Duke University, Durham, NC, USA

The spacing effect in list learning occurs because identical massed items suffer encoding deficits and because spaced items benefit from retrieval and increased time in working memory. Requiring the retrieval of identical items produced a spacing effect for recall and recognition, both for intentional and incidental learning. Not requiring retrieval produced spacing only for intentional learning because intentional learning encourages retrieval. Once-presented words provided baselines for these effects. Next, massed and spaced word pairs were judged for matches on their first three letters, forcing retrieval. The words were not identical, so there was no encoding deficit. Retrieval could and did cause spacing only for the first word of each pair; time in working memory, only for the second.

INTRODUCTION

The spacing effect is one of the oldest (Ebbinghaus, 1885/1964, p.79) and best studied effects in memory research (Crowder, 1976; Dempster, 1988; Glenberg, 1979; Hintzman, 1974; Rubin, 1995). The spacing effect can be of great practical importance because the improvement of spaced over massed practice can be large (e.g. Landauer & Ainslie, 1975) and, unlike other mnemonic aids, it can be implemented without additional study time or training. Nonetheless, its cause remains a mystery. Here we use three mechanisms to account for the advantages of spaced over massed presentations over the short time range used

Requests for reprints should be sent to David C. Rubin, at Department of Experimental Psychology, Duke University, Durham, NC 27708-0086, USA. E-mail: rubin@psych.duke.edu.

This research was supported in part by NSF grant number BNS-9010174. Portions of this research were presented at the 1993 Meeting of the Psychonomics Society. A more detailed version of most of this research is available in K. Braun's (1994) doctoral dissertation, *A retrieval model of the spacing effect*, Duke University, (University Microfilms No. 9500488). We wish to thank J. Zervakis for her help with the figures and G. Craik, R. Crowder, A. Glenberg, R. Greene, & R. Block for their comments.

Kelly Braun is now at Oracle Corporation.

in most laboratory studies. We manipulate these mechanisms during a study phase and test them at the end of the experiment using either an intentional or incidental free recall, cued recall, or recognition test.

The first mechanism is an encoding deficit during the study phase. When an item is repeated consecutively, subjects do not give the second presentation of the item as much attention or effort as they would to a novel item or to the second of two widely separated presentations of the same item. Such study-phase encoding deficits affect later test-phase assessments of memory. This mechanism gains support from studies of spacing (Cuddy & Jacoby, 1982; Hintzman, 1974), overt rehearsal (Ciccone & Brelsford, 1976; Rundus, 1971), self-paced learning (Zimmerman, 1975), and the feeling of knowing in which people tend to overestimate their ability to later recall a massed item compared to a spaced item (Zechmeister & Shaughnessy, 1980). In the laboratory, the subjects do not know at the time of the first presentation whether and when the second presentation will occur. Therefore, the encoding deficit operates mainly on the second of two presentations. In particular, the encoding deficit's effect on the first presentation is limited to processing that occurs after the beginning of the second presentation.

The second mechanism is retrieval during the study phase; in many but not all situations, the second occurrence of an item causes the first to be retrieved. If an item is retrieved from long-term memory instead of being accessed directly from working memory, then its probability of being remembered later during the test phase is assumed to increase. Because the phonological loop of working memory holds only two seconds of material, with the procedures used here a lag of one other item being used as a retrieval cue is, as is argued in more detail later, usually sufficient to clear working memory and require retrieval from long-term memory (Baddeley, 1986).

Support for this mechanism comes from studies of spacing (Greene, 1989; Modigliani & Hedges, 1987; Thios & D'Agostino, 1976; Young & Bellezza, 1982) and from general models of memory (e.g. Bjork & Bjork, 1992). Moreover, study-phase associations can be established between an item and its cues that can aid the cue's effectiveness in a later recall (Jacoby, 1974). Greene provides the most recent argument for the effect of retrieval in spacing. He notes that it should be especially effective in free recall because it encourages the encoding of aspects of the context that differ across conditions as evidenced by an improvement in the temporal ordering of items. In contrast to Greene, we make no claim that retrieval has a differential effect on memory for context as opposed to memory for the item itself.

The third mechanism is the amount of time items spend in working memory during the study phase. This mechanism can work on both the first and the second of two presentations. The longer an item is being used in working memory, the greater the probability of its being remembered later in the test phase. Support for this mechanism comes from a wide host of verbal learning

studies (Atkinson & Shiffrin, 1968; Craik & Watkins, 1973; Jacoby & Bartz, 1972), and from models of prose processing (Kintsch & van Dijk, 1978). Here, time in working memory is varied by manipulating the time an item spends as a retrieval cue while there is a search for a match with an earlier item. It may be possible in theory to collapse much of the encoding deficit factor into time in working memory, but with the techniques used here we have no way of determining this, and so, for now, the encoding deficit is left as a separate factor.

The three study-phase mechanisms used here to account for the spacing effect are not ad hoc. They are all needed to account for data elsewhere in cognitive psychology, and they all have extensive empirical support. Two of the mechanisms, retrieval and time in working memory, are specific and well defined in most cognitive architectures. The third, encoding deficit, is not. It could be caused by any of a host of factors. Because our main focus is on the spacing of once-presented words, where encoding deficit is not a major factor, we will do little to clarify this situation.

In order to integrate these three mechanisms in a model of spacing we need to consider the role of once-presented words in understanding the spacing effect. Some uses of once-presented words in understanding the spacing effect are standard. For instance, a common mechanism proposed to explain the spacing effect is encoding or contextual variability (Glenberg, 1979). According to this mechanism the encoding context varies during the course of an experiment. Items that are spaced as opposed to massed have a broader range of contexts and therefore more and different retrieval cues. We did not include this mechanism because the encoding variability explanation should hold for once-presented as well as twice-presented items, and it does not (Glenberg & Lehmann, 1980; Ross & Landauer, 1978). If two different words are spaced on a list, the probability of at least one of them being recalled should be greater than it would be if they were massed, but is not.

Another use of once-presented words has been to provide two baselines against which spacing can be measured. If the probability of recall of a twice-presented word equals the probability of recall of a once-presented word then the mnemonic costs and benefits of the second presentation of the twice-presented word must just balance. A second baseline can be formed from once-presented words. Ross and Landauer (1978) noted that the probability of recall of a twice-presented word could be compared to the probability of recall of at least one of two once-presented words. If the two presentations of a twice-presented word are independent then the probability that at least one of their presentations is recalled should be equal to the probability that at least one of two different once-presented words is recalled. If there is not independence, then there can be higher recall (Rubin & Wallace, 1989; Watkins & Kerkar, 1985), but the independence assumption is the theoretically relevant baseline. If the two different once-presented words, a and b, are assumed to be independent, the

probability of at least one being recalled is $p(a) + p(b) - p(a)p(b)$. Both $p(a)$ and $p(b)$ are estimated by the average probability of a once-presented word being recalled, p , so this reduces to $2p - p^2$. Recall equal to $2p - p^2$ indicates that the mnemonic costs and benefits of the second presentation of the repeated word equal that of an unrepeated word. For these reasons, once-presented words are included in the first two experiments.

In order to use once-presented words as a baseline, they should be of the same difficulty as the twice-presented words. Here, we approximate this by assigning the same words to different subjects as either a once- or twice-presented word and by minimising serial position effects by placing the once- and twice-presented words in pseudo-random orders in 125-word lists which are surrounded by 10-word buffers.

In addition to these existing uses of once-presented words to calculate baselines against which to measure effects, they can also be used to form massed and spaced pairs. The spacing effect is usually considered with identical items presented twice. But the items need not be identical. To argue otherwise is to argue that the basic mechanisms of cognition change when the same item is repeated. Rather the same theory must account for the spacing effect in identical and different items, although specific mechanisms may be needed for either identical or different items. For instance, encoding deficits should be absent in the incidental learning of different items. In the main methodological contribution of this paper, once-presented words are used to separate the effects of two of the three hypothesised mechanisms. If different but related words are included in a list, they can be made to function as pairs for the subjects. But because they are different, we can attribute any change to either the first or the second member of a pair. Only the first member of the pair can benefit from retrieval. Only the second member of the pair varies the time it spends in working memory as a cue for retrieval as a function of lag. It is possible to infer differences between the first and second presentations in the spacing effect by using the same word if subjects judge the frequency of presentation of modalities and one presentation is auditory and one is visual (Hintzman, Block, & Summers, 1973) or if the words are presented in different colours (Watkins & Kerkar, 1985), but the use of different words allows a simpler more direct test.

In Experiment 3, during the study phase, we have subjects judge whether different words match on their first three letters. This variant of a continuous-recognition task forces the retrieval of the first item if it is not in working memory; that is, if it is spaced. Therefore increases in the recall of the first (but not the second) item with spacing can be attributed to retrieval of that first item. Because the second item spends more time in working memory if it is being used as a probe or cue to search for a spaced as opposed to a massed first item, increases in the recall of the second (but not the first) item with spacing can be attributed to the second item's increased time in working memory. That is, the

use of different words matching on their first three letters allows us to isolate retrieval from time in working memory.

As long as the occurrence of the second item is unpredictable to the subject, encoding deficits can only occur after the second item is presented. The second item can, however, serve to stop processing on the first item, or a mechanism could be postulated that depends on the temporal overlap of the processing of both items (Landauer, 1969). If the two items are different, no encoding deficit should occur under incidental learning conditions because there are no encoding processes that should be reduced. That is, there is no prediction of encoding deficit under mechanisms such as consolidation, habituation, rehearsal, or voluntary attention (Hintzman, 1974). Under intentional learning conditions, however, subjects may choose to reduce controlled processing such as rehearsal, or voluntary attention.

Consider a model of the spacing effect that integrates the three mechanisms of encoding deficit, retrieval, and time in working memory. Start with an incidental memory task during the study phase to make somewhat plausible the assumption that our subjects perform the processes needed to do the cover task we set them, and only those processes. For simplicity, let us start with a continuous recognition study-phase task (like that to be used in Experiment 1) in which the subjects have to decide whether or not the word currently shown on a computer screen has occurred earlier in the task. After some buffer items, the sequence, *shadow, shadow, volume, alcohol, volume*, appears at a four-second rate. For the once-presented word, *alcohol*, there is no encoding deficit because the word has not occurred before. There is no retrieval of this word when it appears because it has not appeared before. There is considerable time spent in working memory as the word is used as a cue to search the earlier list in a vain attempt to find a match. In our procedures it will turn out that deciding correctly that a word did not occur earlier will take about as long as deciding correctly that it did occur if it is the second word of a spaced pair. Both will take about two seconds or about the time needed to clear the phonological loop of working memory. The level of free recall or recognition of this once-presented word during a surprise test phase, which follows the continuous-recognition study phase, provides our baseline level.

For the massed presentation of the item, *shadow*, the first presentation would result in the same effects and therefore the same level of initial learning as the once-presented word. The second presentations would suffer some degree of encoding deficit, would add no retrieval effect because the word that needs to be recognised is still in working memory, and would not benefit much from added time in working memory because the recognition match would occur quickly. Thus, if the encoding deficit were complete for the second presentation, and the added active time in working memory were negligible, the massed presentation would provide equivalent performance to the once-presented word in a later surprise recall or recognition test given at the end of the experiment. Increases

above this level can be attributed to a less than complete encoding deficit of the time spent in working memory.

Consider next the spaced presentation of the word, *volume*, at a lag of one. That is, one word occurs between the two presentations of *volume*. Here we would expect less encoding deficit than the massed condition because another word has intervened. There would be a retrieval of the first presentation of the word, *volume*, because working memory has been cleared of the first presentation of the word, *volume*, by the intervening word, *alcohol*, and the search through long-term memory for a match for that word. Finally, time in working memory needs to be considered. For the first presentation of *volume*, there should be as much time as a once-presented word because a full search without finding a match would be made. For the second presentation of *volume*, there should be more time than for a massed word because *volume* was used as a cue to search long-term memory. Thus the spaced presentation should result in greater recall at the end of the experiment.

The clearing of working memory with one intervening item may seem odd to those used to thinking in terms of a 30-second short-term memory, but given estimates of less than two seconds for storage in the phonological loop of working memory (Baddeley, 1986) and the search and match function being performed, this is a reasonable assumption. Assume *volume* is on the screen and the subject tries to match it with earlier words. In our experiment, for this example, the subject would have just spent four seconds with the word *alcohol* on the screen. We will report later that for approximately the first two of those four seconds the subject would have been searching for earlier matches for *alcohol*. Many words would have been retrieved to attempt the match to *alcohol*, in most conditions enough to clear the phonological loop of working memory of its storage of *volume*. Moreover, clearing of working memory with one intervening item is consistent with the finding in the laboratory spacing literature that there are large differences between lag zero and one (the spacing effect) but small, often non-significant, difference among lags longer than zero (the lag effect) (D'Agostino & DeRemer, 1973; Hintzman, 1974; Toppino & Gracen, 1985; Underwood, 1970; Waugh, 1970).

If the same list were repeated, but the task given the subjects was to judge whether the words had six or seven letters (like that to be used in Experiment 2), then the results of a surprise test at the end of the experiment would be different. If it is as easy to count letters as to remember past presentations, there would be little reason to expect a working-memory or encoding-deficit difference between massed and spaced pairs, and there would be no need for a necessary or effortful retrieval of the first item given the second. That is, from the three mechanisms presented here, no spacing effect is expected and both massed and spaced presentations should be near the $2p - p^2$ level.

If the subjects knew that a recall test was to follow, the results would again be different. Extra processes would occur that would yield higher levels of recall

and in the 6/7-letter task there would be a spacing effect because there might be a voluntary, controlled encoding deficit, and intentional learning processes would certainly involve noting whether a word occurred earlier on the list, thereby reinstating the retrieval and time-in-working-memory process for spaced presentations.

Finally consider what would happen for incidental learning using a modified continuous-recognition task (like that to be used in Experiment 3) in which the second occurrence of *shadow* is changed to *shatter*, the second occurrence of *volume* is changed to *voltage*, and in which subjects are asked to judge whether word pairs match on their first three letters. Because the words are different and because there is no intentional learning, there is no reason to expect an encoding deficit in any condition. Removing any encoding deficit allows us to focus on the two memory-increasing mechanisms. The difference between a massed and a spaced presentation for the first member of the pair of items is that the word from the spaced presentation would benefit from a later retrieval. The difference between a massed and a spaced presentation for the second member of the pair of items is that the word from the spaced presentation would benefit from a longer time in working memory when it was being used as a cue during the search for the retrieval of the first item. Finally, the difference between the first and second items of a massed presentation is that the second item would suffer from less time in working memory as a cue during the search for the retrieval of the first item. As no mechanism has added or subtracted to the first item of the massed pair, it should be comparable to the once-presented words.

If the two words of each pair had been identical, the effects of encoding deficit would return, we would not be able to assign the effects of retrieval to the first member of the pair, and we would not be able to assign the effects of time in working memory to the second member of the pair. Nonetheless, the same theory would account for identical and different words with minor variations caused by the details of the task. In particular, because matching identical words is easier than matching words on their first three letters, there might be slightly smaller differences between massed and spaced pairs in the effects of retrieval and of time in working memory.

The three experiments that follow test this model by creating the situations just described. The model is qualitative in that we have no way to assess, a priori, the relative contributions of each of the three components in general or in the specific procedures used here. It does, however, make many specific, binary predictions because, as just outlined, across various study-phase conditions each of the three components is claimed to have either a measurable effect or no effect on later test-phase memory.

The choice of tasks just illustrated serves not only the purpose of testing the proposed model, but also of excluding other recently proposed mechanisms. In particular, Challis (1993) has argued that deep processing is needed for the spacing effect in cued recall. The tasks used here of determining the number of

letters in a word or judging whether words match exactly or on their first three letters are as close to prototypical shallow tasks as one can get (Craik & Lockhart, 1972). Thus obtaining spacing here would be data inconsistent with the importance of deep processing in the spacing effect. Similarly Greene (1989) has argued that the spacing effect in free recall depends on study-phase retrieval whereas the spacing effect in cued recall and recognition depend on an encoding deficit. Motivated by this distinction we have included free recall and recognition tasks in the first two experiments even though this means that in setting up the experiment we have to carefully balance possible floor effects in recall with possible ceiling effects in recognition.

EXPERIMENT 1

Experiment 1 ensures that we can observe the spacing effect with pairs of identical words when there is forced retrieval of the first member of each pair.

Method

Materials. A set of 75 target words and 20 buffer words were used. The words were nouns with either six or seven letters and a memorability score between .45 and .55 as normed in Rubin and Friendly (1986). The study list had a total of 145 presentations composed of 10 primacy buffer words, 10 recency buffer words, and 125 presentations of target words. These 125 presentations consisted of 50 target words presented twice and 25 target words presented once.

The study list was constructed so that every subject saw the same order of spacings; however, the words at those spacings were varied in three versions of the list. The order of spacings was determined by creating five unique lists of 25 word slots, each containing five once-presented words and two pairs of twice-presented words at each of the five lags (i.e. number-of-intervening-items) conditions. In these 25-word-slot lists the placement of the five different lags (lag 0, 1, 2, 4, and 8) appeared in different serial locations. These five lists were randomly appended to each other to create a 125 word-slot list for target words. To create the final three versions seen by the subjects, words were randomly drawn for the target word list with the stipulation that each word appear in one of the three lists as a single presentation and in the two other lists at different lags. The buffer words remained constant and in the same order in all versions of the study list.

A forced choice recognition test followed the presentation of all the words for half the subjects. Each target was paired with four distractor items. All five choices began with the same letter. The distractor items for the forced choice test were chosen randomly from words in Kucera and Francis (1967) that had the same first letter and frequency as the target word. If there were not enough words with the same frequency, a range of frequencies was used. To determine the range, the frequency of the target word plus and minus one was used first. In

a few cases of very high-frequency target words the rule used for determining the range was from one half to twice the target word's frequency. The few target words that had a frequency of zero in Kucera and Francis were matched with words that had a frequency of one. There were two different random orders of this recognition test.

Design and Procedure. Intentionality of learning (intentional versus incidental) and type of memory test (recall versus recognition) were between-subjects factors. Spacing (massed versus spaced) and number of presentations (once versus twice) were within-subject factors.

The 48 Duke undergraduates, who participated as part of a course research requirement, were assigned to conditions in a revolving order and were tested individually. The experimenter was in the room but not in the subject's view when the subject was watching the screen of an IBM-compatible personal computer. Subjects were presented with one word at a time at the centre of the computer screen using the MEL psychology experiment environment. The word appeared for four seconds during which time the subject had to read the word out loud and respond by pressing the correct key on the computer's keyboard. The response keys were labelled with stickers printed either yes or no. The word remained on the screen for four seconds no matter how quickly the subject responded. Between each word a fixation cross was presented for one second. Subjects were informed that there were words repeated within the list and that presentation of the list would take approximately 10 minutes.

For the continuous-recognition task, subjects were asked to decide if the word on the screen had been presented in the list before. The correct answer was yes for the second occurrence of each of the twice-presented words and no for the first occurrence and for all of the once-presented words. Subjects, who were given a practice list of 10 words, could ask questions about the procedure and the experimenter corrected them if necessary during the practice list. All subjects demonstrated an understanding of the task before they were given the experimental list.

The instructions, which were read by the experimenter, differed according to the intentionality of learning. Subjects to the intentional learning conditions were informed that they would be asked to produce as many words as they could remember after the entire list was presented. Subjects in the incidental-learning condition were not informed of a later test. These subjects thought the word-decision task was the only task. In this experiment as well as all others reported here, subjects in incidental conditions showed surprise when given the retention test.

After the study-phase list, all subjects in the free-recall test-phase condition were given a blank, lined sheet of paper and asked to write down as many words as they could remember from the list. They were told they had 10 minutes to complete the task. However, no subject produced any words after the first five

minutes. Subjects in the recognition test-phase condition were given a multiple-choice test and asked to circle the word they thought was previously presented in the list. They were told to guess if they were not sure.

Scoring of Data. The data collected were the number of target words correctly recalled or recognised. Buffer words recalled were not scored in recall or tested in recognition. For this and all experiments that used continuous recognition, responses were checked to ensure that second occurrences of words were recognised as repetitions. Words not recognised during the study phase were eliminated from analysis during the test phase because it could not be assumed that they produced a retrieval of the first presentation of the word.

Results

Study-phase Performance. Subjects were accurate at the continuous recognition tasks, making a total of 1.8% errors (106 out of 6000 decisions). Of these 106 incorrect responses only 23 were words not correctly identified as repetitions of a previously presented word. These 23 responses (0.3%) were removed from the analysis. No analysis of errors was done on these few errors.

Spacing Effects. In this experiment and all that follow, the term *massed* is used for zero lag and *spaced* is used to mean the average of all non-zero lags. Here and in the literature a robust massed versus spaced (zero lag versus non-zero lags) difference is often observed without a consistent linear trend or significant differences among the non-zero lag conditions. The results of such lag, as opposed to spacing, effects will be considered after all the experiments are presented.

Means are shown in Fig. 1. A 2 (spacing: massed versus spaced) \times 2 (intentionality: intentional versus incidental learning) \times 2 (test type: recall versus recognition) ANOVA was performed on the twice-presented words. Statistics are based on percent correct. Unless noted as not significant (ns), all results throughout the paper are statistically significant at $P < .05$. The main effects were: spacing $F(1,44) = 18.44$, $MS_e = 67.93$, intentionality $F(1,44) = 2.46$, $MS_e = 208.36$, ns, and test type $F(1,44) = 664.87$, $MS_e = 208.36$. No interactions were significant—all $F(1,44) s < 1.59$. The lack of an intentionality effect may be due to a ceiling effect in the recognition data. In hindsight we were too ambitious when we contrasted free recall and recognition in the same study. This forced us to have either close to ceiling or close to floor effects on one of the measures, even though we made the recognition task as hard as we could. Although these ceiling effects may have removed the intentionality effect, the reader should note that we observed spacing effects even where this problem was at its worst. Moreover, with respect to spacing, we had large main effects with no interactions, which is not a classic ceiling effect pattern. Thus, the

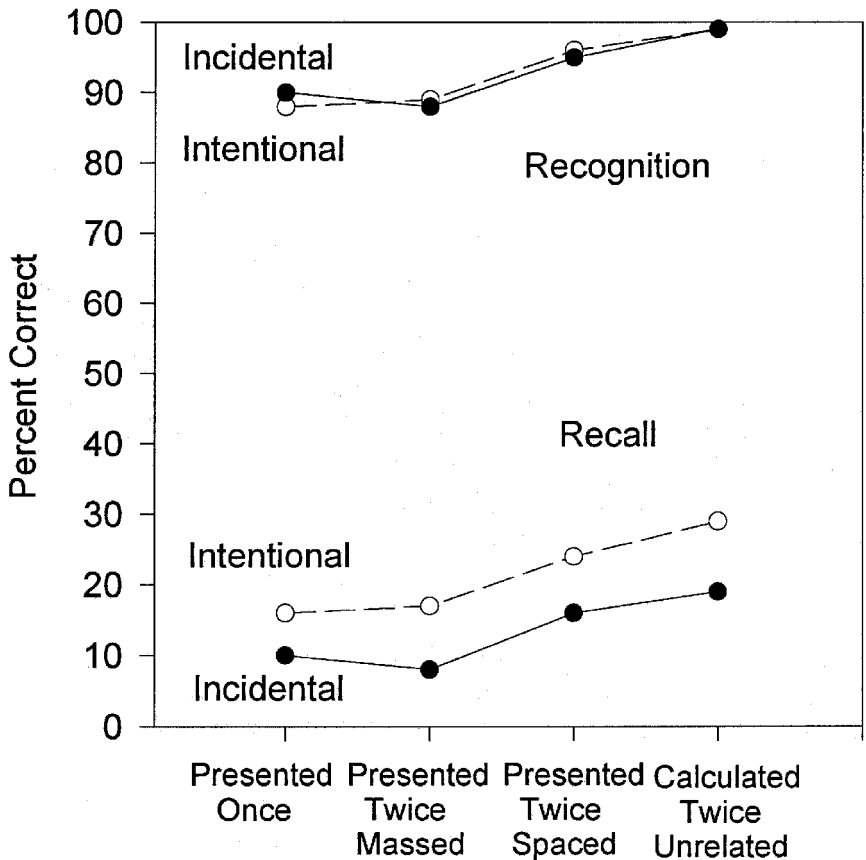


FIG. 1. Percentage of words free recalled or recognised on the final test in Experiment 1 as a function of whether a word was presented once, twice massed, or twice spaced, or the percentage that at least one of two unrelated words would be remembered was calculated (calculated twice unrelated). The subjects, who judged whether the word being shown had occurred earlier in the list, were either aware (intentional) or unaware (incidental) that a memory test would follow while making the judgements.

effortful retrieval forced by the study-phase continuous-recognition task appears to be sufficient to induce a spacing effect independent of the intentionality of learning.

Single versus Twice-presented Words. Also shown in Fig. 1 are the probabilities of recall and recognition of once-presented words and the probabilities of at least one of two once-presented words being recalled, as calculated from $2p - p^2$. There is not a significant difference between the

probability of remembering a once-presented word and a massed twice-presented word, $F(1,44)=0.03$, $MS_e=98.13$, there is an effect of test type, $F(1,44)=888.03$, $MS_e=156.62$, but not an interaction, $F(1,44)=0.00$, $MS_e=98.13$. In addition, the value of $2p - p^2$ is greater than the spaced twice-presented words, $F(1,44)=4.36$, $MS_e=49.30$, there is an effect of test type, $F(1,44)=940.70$, $MS_e=143.90$, but not an interaction, $F(1,44)=0.17$, $MS_e=49.30$.

These results imply that the probability of at least one of two different words being recalled is greater than the probability of a twice-presented word being recalled, and if the twice-presented word has a massed presentation its probability of being recalled drops to approximately what it would be with only one presentation. Thus, overall spacing does not increase the probability of recall above that of recalling at least one of two different words, but does help to minimise the deficit. Compared to two different words, the massed presentation of identical words suffers both an encoding deficit and also less time in working memory because it does not take as long for the second presentation to be judged in the continuous recognition task. For the spaced presentation of identical words, both the detrimental effects of the encoding deficit and time in working memory are less, because the intervening items require an increase in encoding and a larger search time to determine that the second presentation was a repetition. In addition there is a benefit to the first presentation from retrieval.

EXPERIMENT 2

Experiment 2 is the same as Experiment 1, except that the continuous-recognition study-phase task is replaced by a 6/7-letter judgement task. All other aspects of the design are identical. The continuous-recognition task forced the effortful retrieval of the first presentation of each twice-presented word at the time of the word's second presentation. The 6/7-letter judgement task does not, but such a retrieval would be expected if the subjects knew that a later recall were expected. Thus, as predicted by the model, the second experiment tests whether the spacing effect occurs with intentional but not incidental learning using the 6/7-letter task and that the same effects hold for both recall and recognition.

Method

The identical materials, design, and procedure were repeated except that the continuous-recognition judgement was replaced with a 6/7-letter task and that 48 Duke undergraduates who did not take part in Experiment 1 were recruited. One half of the subjects were asked to decide if each word contained six letters and the other half decided if each word contained seven letters.

Results

Study-phase Performance. Subjects were accurate at the letter judgement task making only 2.9% errors (174 errors out of 6000 decisions). These items were not included in further analyses.

Spacing Effects. Means for these main effects and interactions are shown in Fig. 2. A 2 (spacing: massed versus spaced) \times 2 (intentionality: intentional versus incidental learning) \times 2 (test type: recall versus recognition) ANOVA was performed on the twice-presented items. The main effects were: the not-significant effect of spacing, $F(1,44) = 1.85$, $MS_e = 74.66$; intentionality $F(1,44) = 9.00$, $MS_e = 149.31$; and test type $F(1,44) = 392.31$, $MS_e = 149.31$. The 2-way interaction of spacing \times intentionality $F(1,44) = 5.04$, $MS_e = 74.66$, was the only significant interaction—all other $F(1,44)s < 1.69$.

An examination of Fig. 2 indicates that the lack of a main effect for spacing and the spacing by intentionality interaction appear to be caused by the predicted outcome of no spacing in the incidental condition which did not require or encourage the retrieval of the first presentation of the twice-presented items. In fact, the difference between the massed and spaced incidental conditions is in the wrong direction for a spacing effect. To test this further, the data from the intentional and incidental conditions were analysed separately. For the intentional data there are significant spacing and test-type effects, $F(1,22) = 5.28$, $MS_e = 91.84$ and $F(1,22) = 150.66$, $MS_e = 355.10$, respectively, but no interaction, $F(1,22) = 1.54$, $MS_e = 91.84$. For the incidental data there is a significant test-type effect, $F(1,22) = 276.71$, $MS_e = 200.76$, but not a significant spacing effect $F(1,22) = 0.51$, $MS_e = 57.47$ or interaction, $F(1,22) = 0.27$, $MS_e = 57.47$.

The lack of a spacing effect in the incidental-learning recall condition could be considered as due to a floor effect, but this is unlikely to be the reason. First, similar levels of recall produced spacing effects in Experiment 1. Second, recognition also failed to show a spacing effect in incidental learning and its levels are far from floor or ceiling effects. Nonetheless, in the remaining experiment we switch from free to cued recall to bring the levels of recall and recognition closer together in order to avoid possible floor and ceiling effects. Similarly, it cannot be a lack of power that led to the lack of a spacing effect in the incidental learning recall and recognition conditions. First, the means are not in the correct direction. Second, there is a spacing effect for the intentional condition. Third, the overall spacing effect in Experiment 1, which differed only in the study-phase cover task and thus which had the identical design and number of subjects, was $F(1,44) = 18.44$, $P < .0001$. Thus, there is certainly power enough in the design to detect spacing at the .05 level.

Single versus Twice-presented Words. Figure 2 also shows the probability of recall and recognition of once-presented words and the probability of at least

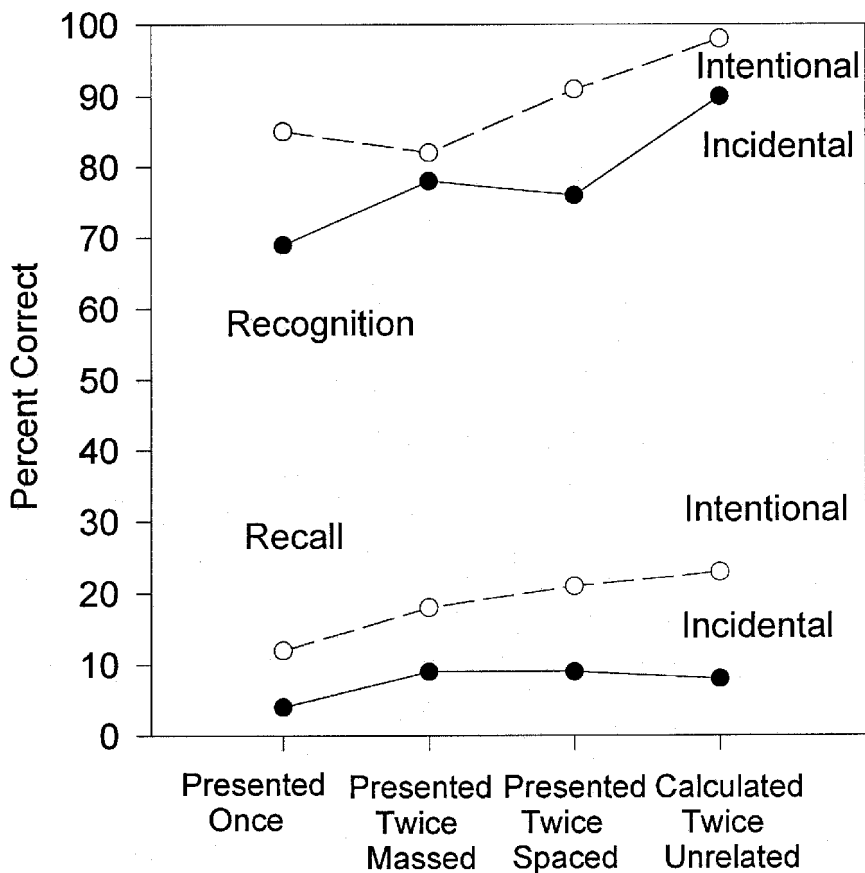


FIG. 2. Percentage of words free recalled or recognised on the final test in Experiment 2 as a function of whether a word was presented once, twice massed, or twice spaced, or the percentage that at least one of two unrelated words would be remembered was calculated (calculated twice unrelated). The subjects, who judged whether words had six or seven letters, were either aware (intentional) or unaware (incidental) that a memory test would follow while making the judgements.

one of two once-presented words being recalled, as calculated from $2p - p^2$. For the incidental data, in which there is no spacing effect, massed pairs are recalled more than once-presented words, $F(1,22) = 12.65$, $MS_e = 49.87$, there is an effect of test type, $F(1,22) = 257.36$, $MS_e = 207.75$, but not an interaction, $F(1,22) = 1.41$, $MS_e = 49.87$. In addition, the value of $2p - p^2$ is greater than the special twice-presented words [spaced versus $2p - p^2$, $F(1,22) = 15.17$, $MS_e = 33.14$, and there is an effect of test type, $F(1,22) = 637.99$, $MS_e = 102.34$, and their interaction, $F(1,22) = 17.18$, $MS_e = 33.14$]. To interpret their interaction, separate analyses were done for recall and recognition. For recall there is

no difference between the value of $2p - p^2$ and the spaced twice-presented words, $F(1,11) = 0.06$, $MS_e = 16.51$, whereas for recognition there is a difference, $F(1,11) = 21.52$, $MS_e = 49.77$.

If the 6/7-letter judgement task in the incidental learning condition is done by judging each word separately then the recall level of a massed presentation, a spaced presentation, and at least one of two different words should all be the same. Moreover, the recall level of a massed presentation should be greater than that of a single presentation. This occurs for recall and for recognition to the extent that, unlike Experiment 1, the massed presentation is greater than that of a single presentation. The difference between the spaced presentation and the $2p - p^2$ calculation for recognition is the only discrepancy from the predicted results.

For the intentional data, a different pattern emerges. There is not a significant difference between the probability of recall of a once-presented word and a massed twice-presented word, $F(1,22) = 0.29$, $MS_e = 92.21$, there is an effect of test type, $F(1,22) = 155.58$, $MS_e = 361.91$, but not an interaction, $F(1,22) = 3.47$, $MS_e = 92.21$. In addition, the value of $2p - p^2$ is not greater than the spaced twice-presented words [spaced versus $2p - p^2$, $F(1,22) = 1.13$, $MS_e = 68.25$, test type, $F(1,22) = 442.68$, $MS_e = 142.74$, and their interaction, $F(1,22) = 0.98$, $MS_e = 68.25$].

Because intentional learning involves all the aspects of the incidental learning condition plus an attempt to learn the words, there should be the same effects as were present in the incidental condition mixed with the effects of trying to learn the words. The latter effects can be approximated by the continuous recognition task of Experiment 1, because part of the effort to learn would involve attempts to note whether a word had occurred earlier in the list. The intentional learning results are much closer to those of the continuous-recognition results of Experiment 1, the only difference being that there the $2p - p^2$ value is not greater than that of the spaced presentation.

The overall pattern of results should occur under the model specified in the introduction: (a) the deficit in the encoding of massed twice-presented words should not occur in the incidental learning conditions because the words were being processed only to count their letters, but should occur in the intentional learning conditions where added mnemonic processing took place; (b) there should be no effortful retrieval in the incidental conditions for the first presentation of the spaced twice-presented words because there was no need to retrieve it under the cover task, but there should be effortful retrieval in the intentional conditions because the mnemonic processing required it. The one exception to this conclusion is the $2p - p^2$ calculated value for the incidental-learning recognition subjects, which is too high. Thus as in Experiment 1, this overall pattern of results is what would be predicted from the proposed model of the spacing effect.

EXPERIMENT 3

Having demonstrated the spacing effect when effortful retrieval of the first member of a pair occurred and the lack of a spacing effect when that retrieval did not occur, we search for the precise locus of the effect. Experiment 3 uses pairs in which the first and second members differ. In this way the locus of the spacing effect can be attributed to one or the other member. This novel manipulation is useful in the procedures used here because study-phase retrieval can occur only for the first member of a pair and only when it is retrieved, because it is no longer in working memory when it needs to be compared to the second member of the pair. In contrast, differences in time in working memory with spacing can occur only for the second member of a pair because the search time for using the second item as a cue increases with lag but the search time for using the first item as a cue does not. Encoding deficit is not a factor for different items. In intentional learning there might be encoding differences if the subjects thought they had learned the massed pairs better, but in the incidental-learning continuous-recognition task used here such differences are not likely.

The strategies our subjects used to increase their intentional learning above that of incidental learning are difficult to uncover. We therefore restrict our study-phase procedures to incidental learning with a demanding cover task that provides more specific indications of processing. Moreover, in order to force and be able to monitor retrieval of the first presentation of a pair during the study phase, we use only variants of the continuous-recognition cover task from Experiment 1.

If retrieval contributes to the spacing effect then it should be possible to produce the spacing effect using items that cue retrieval even if the cueing item is not an exact repetition. Experiment 3 was designed to test this hypothesis. The task for the subject is to decide if they have seen a word previously presented in the list that matches on at least the first three letters. In order for the subject to correctly respond "yes", they should retrieve the previous matching occurrence. In contrast to the predictions made here, previous studies have suggested that only exact repetitions produce the spacing effect. For instance, Dellarosa and Bourne (1985) found the spacing effect with exactly repeated sentences but not "gist" sentences, and Glanzer and Duarte (1971) found an attenuated spacing effect when language was changed in bilingual subjects. The difference here is that we force the retrieval of the exact earlier item even though the exact earlier item itself was not repeated.

In our first work using this procedure some subjects used a strategy in which they only looked at the first three letters of the word instead of reading the entire word. To eliminate the use of this strategy, we added a few unscored cases of identically repeated pairs of words as well as three-letter combinations that occurred three times. The subjects were asked to choose among three alternatives: whether the word being shown (1) matched no earlier words, (2)

matched an earlier word verbatim, or (3) matched an earlier word only on the first three letters. The primacy buffer contained examples of all three alternatives so that subjects were alerted from the beginning of the experiment that using a strategy of not reading the whole word would not be useful.

In an attempt to raise recall percentages, a cued-recall task was used instead of the free-recall task of Experiments 1 and 2. The cued-recall test prompted with the first three letters of each target word: the letters for which a match would occur. There was space for two words to be written down and subjects were informed in the instructions to write two words if they could remember seeing two different words. Because there should be no encoding-deficit effect with the different-word pairs, we expect the spacing effect to be smaller. To increase the power of the experiment only different-word pairs are used, thereby increasing their number.

Methods

Materials. The study list was constructed of 75 pairs of target words and 10 primacy and 10 recency buffer words. Each target pair begins with the same three letters and this three-letter combination is unique to each pair. The words were drawn from the same pool as used in the previous experiment. However, because subjects were looking for words that matched on the first three letters, approximately one third of the words in the original list had to be replaced. This was to ensure that only one word could be retrieved as a yes response. Replacement words were rated between .40 and .58 on memorability in Rubin and Friendly (1986). For each of these nouns, a second six- or seven-letter noun matching on only the first three letters was found in Kucera and Francis (1967). Obvious associations were avoided. Seven additional words were needed as once-presented filler items to produce the needed spacings of the list. All buffer and filler words began with a unique three-letter combination different from any of the target words.

The word list was constructed so that there were 15 pairs of words at each of the five spacings. This made for a total of 75 word pairs. To ensure that the different spacings were presented evenly over the list, five intermediate word-slot lists were created. These intermediate lists were created so that the least number of filler items were needed to properly space the items. These five lists were then appended to each other. In some cases it was possible to remove filler items by starting the next intermediate list within the last intermediate list without disrupting the spacing. The final word-slot list consisted of the 75 target pairs and seven filler items.

Three lists were created by taking the final word-slot list and adding the primacy and recency buffers. The buffer and filler words were the same for all versions of the list. Next each pair of words was randomly assigned to a different spacing and a different area within each of the three lists. Finally for each of the

three lists another list was created by switching the position of the first- and second-presented items of the pair. These manipulations created six different presentation versions of the 177-word list. Next, manipulations were made to encourage the subjects to examine each whole word and not just its first three letters. Each of the six study lists was reduced to 65 instead of 75 pairs of target words with 13 instead of 15 pairs at each spacing. The 20 words removed were replaced to provide five unscored triplets of identically repeated words and five unscored triplets of words beginning with the same three letters. Words used to create the triplets were chosen from Kucera and Francis (1967). To accommodate the triplets, two pairs of each spacing were chosen and used as part of a triplet. The position of these replaced pairs were chosen to take advantage of filler items and primacy buffer words to make triplets. Because the spacing positions were the same for all six versions of the study list, three different groups of words were eliminated.

The multiple-choice recognition test had three foils per target word. The foils for each target item did not start with the same letter as the correct response—however, there were several foils throughout the test that did. As in the earlier experiments, the recognition test was given in two different orders.

Procedure. The procedure was identical to the incidental-learning conditions of Experiment 1 except as noted here, with 48 Duke undergraduates who had not taken part in the earlier experiments participating as part of a course research requirement. Subjects did not read the word aloud because they were tested in small groups of one to four. Subjects were asked to respond to each word by pressing one of three, as opposed to two, marked keys on the computer keyboard corresponding to the category to which the word belonged. The three categories were defined as follows: a word that was previously presented in the list; a word that matched on at least the first three letters but was not an exact repetition of a previously presented word; or a word that did not match or matched on only the first one or two letters with a previously presented word.

Scoring of Data. As in Experiment 1, any word not correctly identified as a repetition was removed from the analysis.

Results

Study-phase Performance. A total of 607 out of 3120 (19.5%) second-position words were not correctly identified as repetitions.

Spacing Effects. A 2 (spacing: spaced versus massed) \times 2 (word position: first versus second word) \times 2 (test type: recall versus recognition) ANOVA was performed. All main effects were significant: spacing $F(1,46) = 53.38$, $MS_e = 112.04$, word position $F(1,46) = 33.73$, $MS_e = 77.62$, and test type

$F(1,46) = 290.39$, $MS_e = 218.03$. The spacing \times test type interaction was also significant $F(1,46) = 4.93$, $MS_e = 112.04$, but the other interactions were not—all $F(1,46)s < 1.33$. The means for these effects are shown in Fig. 3.

Because the spacing \times test type interaction was significant further analysis was performed. Two 2 (word position) \times 2 (spacing) analyses were done, one for each test type to ensure that the effect held for each individually. In both the recall and recognition analyses, spacing and word position were significant [spacing was $F(1,23) = 32.73$, $MS_e = 155.32$ in recognition and $F(1,23) = 21.08$, $MS_e = 68.75$ in recall; word position was $F(1,23) = 9.17$, $MS_e = 91.72$ in recognition and $F(1,23) = 26.61$, $MS_e = 65.5$ in recall], and there were no

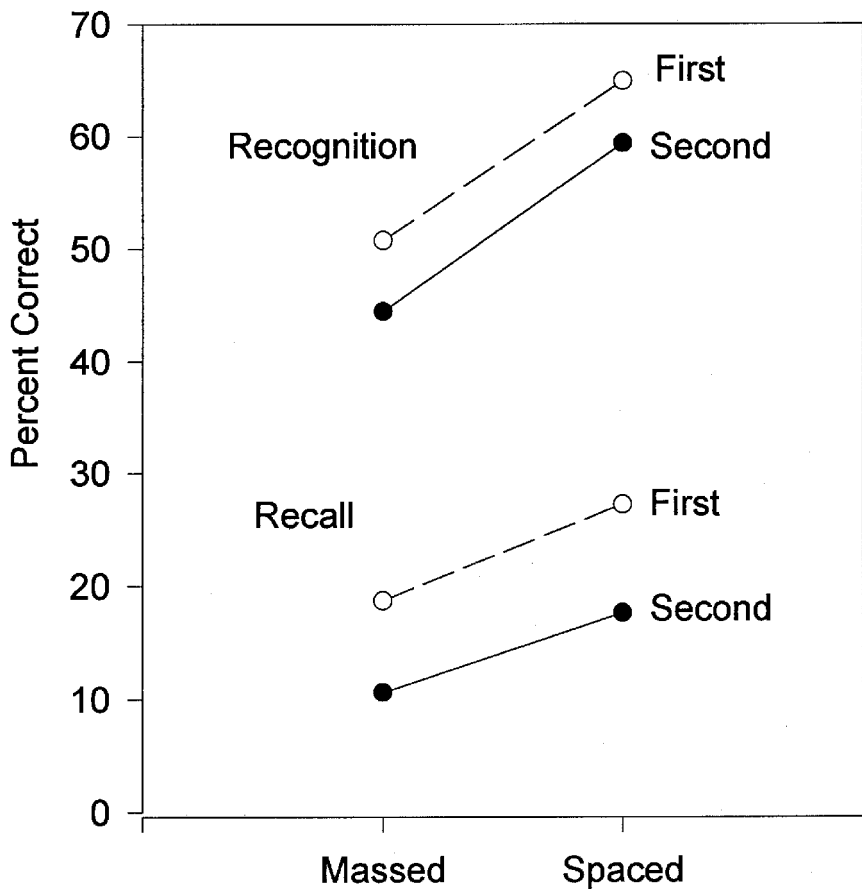


FIG. 3. Percentage of first and second words of pairs that were remembered on a surprise memory test in Experiment 4. Subjects judged whether the word being shown matched an earlier word exactly, on its first three letters, or at most on its first two letters.

significant interactions—both $F(1,23)s \leq 0.14$. Thus, although the spacing \times test-type interaction shows that the effect of spacing is greater in recognition, the effect of spacing is significant for both recognition and recall.

Little can be made of this difference in effect size. Although there is a 14% increase from massed to spaced presentation in recognition compared to 8% in recall, the corresponding percentage increases over the massed baseline are in the other direction with 29% for recognition and 57% for recall. That is, interpretation of the differences in the effect size depends on scaling factors and these have not been set by the mechanisms proposed here.

Analysis of Error Data. Because the second presentation of the pair was often not recognised as a three-letter repetition, there were enough words removed from the main analysis to allow errors to be examined in more detail. No distinction of test type is made because subjects were not aware of the test at the time of the modified continuous-recognition task.

Figure 4 shows a measure of difficulty of retrieval as a function of lag: the proportion of continuous-recognition errors on the second word of each pair out of a possible 13 at each lag for each of the 48 subjects. The errors are caused by a failure to classify the second word correctly as a three-letter repetition, but the fault cannot be assigned solely to the second word of the pair because the measure is a failure to note a match between two words. The general shape of the curve is consistent with earlier continuous-recognition results (see Rubin & Wenzel, 1996, for a review). A more detailed analysis is relevant here.

The overall ANOVA for the proportion of errors is $F(4,44) = 26.73$, $MS_e = 2.30$. On a set of contrasts between lag 0 versus 1, 1 versus 2, 2 versus 4, and 4 versus 8 performed to compare the proportion of error at each lag, only the proportion of errors at lag 4 and lag 8 were not significantly different ($P = .11$) [$F(1,47) = 47.02$, $MS_e = 2.63$, $F(1,47) = 6.76$, $MS_e = 3.78$, $F(1,47) = 4.30$, $MS_e = 3.53$, and $F(1,47) = 2.73$, $MS_e = 4.77$, respectively]. The increase in error rate is evidence that the modified-continuous-recognition task becomes more difficult as a function of lag, especially between lag 0 and 1. As both the percentage of errors and the time-in-working-memory increase to be discussed next show the same pattern, this result also ensures that the time-in-working-memory increase is not due to a speed-accuracy trade-off. However, the size of the effect, in this but not in the earlier experiments, raises concerns that the observed spacing effect is caused by the elimination of items that are difficult to match on three letters in the study-phase, continuous-recognition task. These items, which are eliminated from scoring, may also have been more difficult to remember in test-phase cued recall and recognition tests. The data for all five lags, which are presented after the last experiment, do not support this concern. Remembering does not increase continuously as it should under this hypothesis and the results shown in Fig. 4, but is a step function with lag 0 lower than lags 1–8 which are nearly equal. That is, if study-phase selection effects

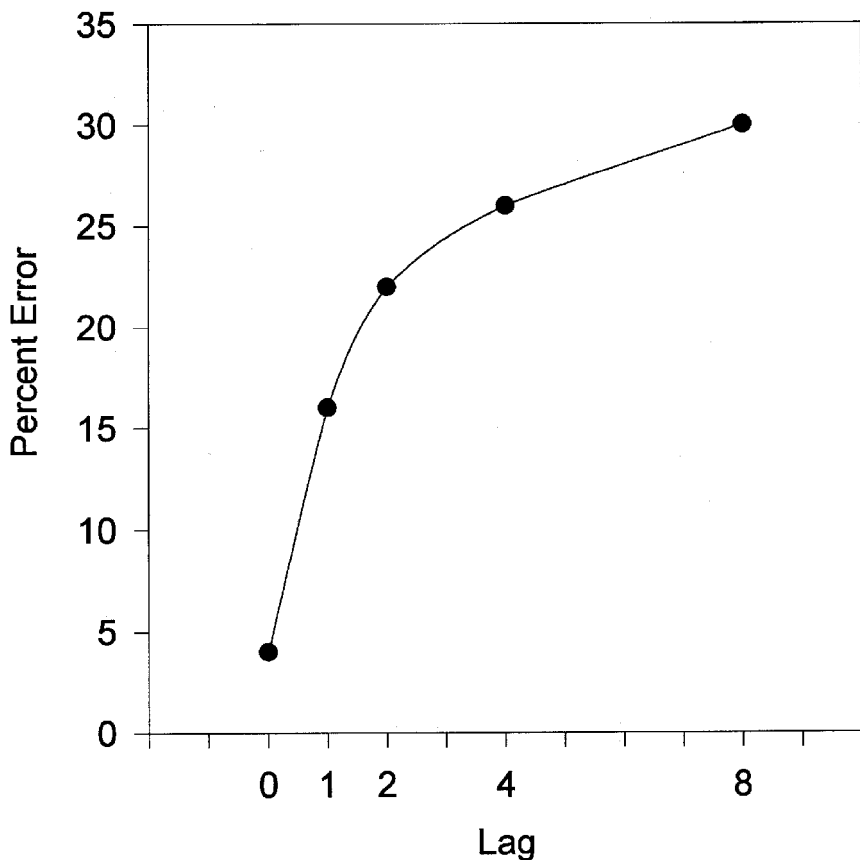


FIG. 4. Percentage of errors as a function of the number of intervening items in the modified continuous-recognition task of Experiment 4.

caused the final test effects, then following the magnitude of study-phase selection effects shown in Fig. 4 the differences in the final test between lag 0 and lag 1 would be approximately equal to the final test differences between lag 1 and lag 8. As shown in the later section on lag effects, they are not.

Next the pairs that were errors of continuous-recognition and that had been removed from the main data analysis were examined for later retention. There were too few errors at the massed presentation (only 23) to include a spacing effect, but a 2 (word position—first versus second) \times 2 (test type) ANOVA was performed to test if there were differences in the percentage of later-remembered words. The means for recall were 1.58 for the first presentation and 1.54 for the second presentation. The means for recognition were 6.00 for the first

presentation and 5.79 for the second presentation. There was no word position effect, $F(1,46) = 0.14$, $MS_e = 2.68$, or interaction, $F(1,46) = 0.06$, $MS_e = 2.68$, however the main effect of test type significant, $F(1,46) = 51.28$, $MS_e = 8.79$. The lack of a significant difference between the first and second word positions on later retention was consistent with the words being treated as single presentations under the model being tested.

Time in Working Memory. Because Experiment 3 demonstrates a spacing effect for both the first and second members of pairs, it allow us to test the time-in-working-memory component of the model. The time spent in working memory by the first member should be long in a futile attempt to find an earlier match. The time spent by the second member should be a function of lag, with massed presentations taking considerably less time than other lags because retrieval is not needed. Because the data reported are for the spacing effect, no main analysis was made among the non-zero lags, but if the time-in-working-memory measure is to have face validity there should be an increase in time with the increase in lag (i.e. with the increase in the number of intervening words and time between the first and second members of the pair).

Figure 5 contains the mean times for the subjects to decide whether a word occurred earlier, matched on its first three letters, or did not match. As with the error data there is no distinction between recall and recognition data because the subjects were not aware of this task at the time of the continuous-recognition task. The data follow the predictions and also indicate that subjects spend approximately as much time deciding that the first member of a pair did not occur as deciding the second member from lag 2, 4, or 8 did. Using seconds as a measure of reaction time, there was a main effect of first versus second, $F(1,47) = 11.82$, $MS_e = 0.19$, of lag, $F(4,44) = 32.91$, $MS_e = 0.04$, and an interaction, $F(4,44) = 20.33$, $MS_e = 0.05$. Considering the first member of the pair's data separately there was no lag effect, $F(4,44) = 1.25$, $MS_e = 0.04$. Considering the second member data separately there was a lag effect, $F(4,44) = 46.32$, $MS_e = 0.05$. A planned comparison between successive time intervals for the second member showed that lag 0 and lag 1 differed, $F(1,47) = 75.44$, $MS_e = 0.09$, but that lag 1 versus lag 2, lag 2 versus lag 4, and lag 4 versus lag 8 did not, $F(1,47) = 2.12$, $MS_e = 0.09$, $F(1,47) = 2.70$, $MS_e = 0.08$, and $F(1,47) = 2.86$, $MS_e = 0.11$, respectively. We cannot claim that these differences beyond lag 1 would not be significant if more subjects were tested or different statistical tests were performed. However, consistent with our claims about the effects of time in working memory during the study phase on later test-phase remembering, these differences beyond lag 1 are certainly smaller than the differences between lag 0 and lag 1.

For the second items, the difference between massed and lag-1 items was on the order of 375ms whereas the average difference calculated from the slope of lags 1-8 is on the order of 30ms/lag. This finding supports the time-in-working-

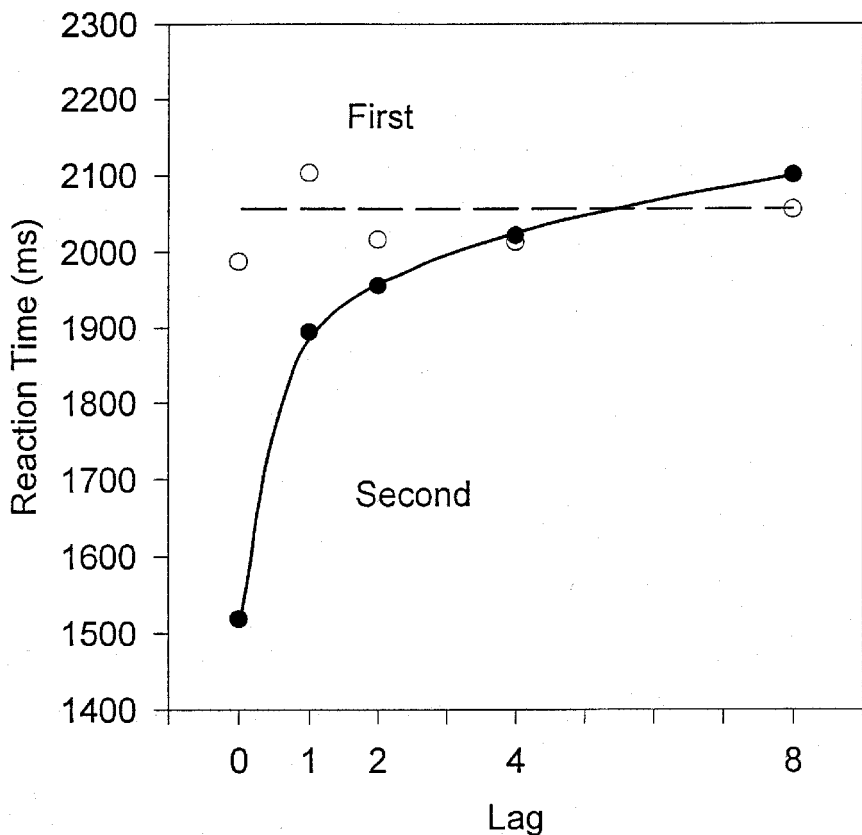


FIG. 5. Reaction time as a function of the number of intervening items in the modified continuous-recognition task of Experiment 4 for the first and second words of pairs.

memory mechanism in that there is typically a much larger difference in later remembering between lag 0 and lag 1 (the spacing effect) than among other lags (the lag effect). It also supports the need for retrieval, or some other item-based processing mechanism, that distinguishes the large effect of the first, as opposed to later, intervening items. The finding, however, is not consistent with the lack of a lag effect reported in the next section. Items at lag 8 spend more time in working memory in the study phase than items at lag 1, but show no increase in their final recall or recognition test.

LAG EFFECTS

As in many other laboratory studies of spacing, there were few effects of lags beyond lag 1. To show this in some detail, Table 1 contains the means for each

TABLE 1
Mean Percentage of Remembering as a Function of Lag

<i>Experiment</i>	<i>Lag</i>				
	<i>0</i>	<i>1</i>	<i>2</i>	<i>4</i>	<i>8</i>
<i>Combined</i>					
1	51	58	56	59	59
2	47	50	47	50	51
3	31	42	41	43	44
<i>Recall</i>					
1	13	20	18	22	21
2	14	17	11	16	16
3	15	23	19	22	26
<i>Recognition</i>					
1	89	96	94	97	96
2	80	83	83	83	86
3	48	61	62	64	61

experiment as well as for recall and recognition separately, as this is the only division that holds over all experiments. Once the zero lag is removed, the differences among the remaining non-zero lags show no reliable trends. For Experiments 1, 2, and 3 shown in Table 1, the main effects for the non-zero lags are $F(3,42) = 1.39$, $MS_e = 89.25$; $F(3,42) = 1.61$, $MS_e = 103.95$; and $F(3,44) = 2.65$, $MS_e = 180.28$, respectively. Moreover, none of the interactions of the non-zero lags with any other factors was significant, even though all experiments retained the significant main effects and interactions involving test type, intentionality, and word position that they had in the earlier analyses which included lag-zero data. This finding is a massive replication of a lag null effect, often in the presence of large spacing effects. It is consistent with the proposed difference between no retrieval at lag 0 and retrieval at all other lags and the small gradations of time in working memory beyond lag 1.

GENERAL DISCUSSION

A model has been proposed borrowing both ideas and evidence from the long history of the laboratory study of the spacing effect, and adding ideas and evidence based on using once-presented words not only as a baseline but also as members of massed and spaced pairs. Here by way of summary, we present the model in full, attempting to enumerate all assumptions about encoding deficit (ED1 – ED3), retrieval (R1 – R6), and time in working memory (WM1), and list areas of ignorance as we proceed.

When an item is repeated exactly in the massed condition, an encoding deficit for the second presentation exists (ED1). We remain neutral on whether

continuing processing or consolidation of the first member of the pair is also shortened and whether the mechanisms for the encoding deficit are mostly controlled or not. For related but not identical items there is no reason to expect an encoding deficit for incidental learning (ED2), although there may be an encoding deficit caused by conscious (i.e. controlled) mechanisms with intentional learning (ED3).

If an item is retrieved then its probability of later being remembered is increased (R1). Retrieval will not occur if the item is in working memory (R2), but for most experimental conditions a lag of one intervening item is sufficient to clear working memory and thus to force retrieval if it is needed for the subject's task (R3). Retrieval is assumed to occur without the need for conscious control if an item is repeated and its earlier presentations are not all in working memory (R4). This assumption is needed to account for differences in the judgement of spacing between exact and different items (Hintzman, Summers, & Block, 1975). In addition to such automatic retrieval, effortful or controlled retrieval can be forced by a cover task, such as continuous recognition, or encouraged by intentional-learning instructions and discouraged by a cover task that focuses on individual items, such as judging length in letters, when incidental learning is used (R5). The results of our incidental versus intention experiment judging six versus seven letters indicate that study-phase effortful retrieval has a much greater effect on later test-phase effortful retrieval (R6).

Most novel for the study of spacing, but well documented elsewhere in the memory literature, time in working memory increases the probability of later test-phase remembering (WM1). One challenge to this assumption in the current study is the lack of a lag effect in the presence of an increase in reaction time (and therefore time in working memory) with increased lag. One problem that remains is that study-phase time in working memory increases beyond a lag of 1, yet we observe no increase in test-phase performance beyond lag 1. Although the time difference between lag 0 and lag 1 is twice that of the difference between lag 1 and lag 8, the theory would still expect some difference. In addition, a distinction may have to be made between active processing and maintenance rehearsal. Although maintenance rehearsal can result in increases in the probability of overall remembering when measured by recognition (Glenberg, Smith, & Green, 1977), we do not know whether it can affect spacing of identical words as measured by either recall or recognition, although this would be of theoretical interest because existing results suggest that it would have little effect on pairs made of related but different items (Glenberg & Adams, 1978).

This set of assumptions, most of which have strong and varied support from the literature, are able to account for the spacing effect data obtained here and for much of the existing literature. Using both identical and different items for the first and second presentations over the course of the experiments we were able to demonstrate the effects of three mechanisms: an encoding deficit, retrieval, and time in working memory. The use of pairs of different items in

Experiment 3 allowed us to isolate retrieval from time in working memory. The three mechanisms could be applied in a straightforward manner.

We could not predict the relative importance of the three mechanisms based on our own model in general terms, but for the particular set of experiments performed, the model allows us to estimate their importance empirically. From the parallel lines between the first and second members of the pairs for both recall and recognition in Fig. 3, it follows from the model that the test-phase mnemonic benefits of study-phase retrieval and added time in working memory are equivalent in Experiment 3. Measuring the effect of encoding deficit is more difficult because it occurs in Experiments 1 and 2 as one of three factors and because it is not well defined as a mechanism and may include time in working memory as a component. Nevertheless, by using Experiment 1, which also has a continuous recognition task and which shows a spacing effect, we can make an estimate. In Experiment 1, once-presented words and massed pairs produce equivalent test-phase performance. As there is no retrieval in the massed pairs, the encoding deficit for the massed items must equal any benefits of the short time the second word of the massed pair spends in working memory. In Experiment 1, the spaced pairs produce lower test-phase performance than would two separate words. Thus there is an encoding deficit for spaced items, which according to the model is less than the encoding deficit for massed items. In recall where there is no ceiling effect, the encoding deficit for spaced items is about half the size of the spacing effect. If we assume that the retrieval and working memory components of the spacing effect have equal effects as they did in Experiment 3, then the encoding deficit for spaced items has the same size of test-phase effect as either retrieval or working memory.

The estimates in the previous paragraph are both rough and dependent on the particulars of the experiment. With added parametric work, however, realistic assessment of the relative and absolute mnemonic benefits and cost of each of the three mechanisms could be made within the model.

In general, spacing was obtained only where it was predicted to occur in the first two experiments in spite of having to balance possible floor and ceiling effects introduced by our inclusion of free recall and recognition in the same study. The methodological problems were minor and when the set of three experiments is considered as a whole, they provide a picture of spaced practice that is consistent with the theory guiding the work. Although our series of experiments can only begin to test the completeness and limits of the model, it is clear that including once-presented items both in baselines and in massed and spaced presentations has clear advantages for theory development.

The question remains of what the laboratory task of spacing has to do with the powerful effects of spacing outside the laboratory, effects that in laboratory terms are lag effects rather than spacing effects (e.g. Bahrck, 1979; Bahrck & Hall, 1991; Bahrck & Phelps, 1987; Landauer & Ross, 1977; Landauer & Bjork, 1978; Rubin, 1995). That is, the long-term spacing effect is not between

massed (lag 0) and spaced (lags 1 and greater) but among lags greater than one. There are many differences between the studies of the long-term spacing effect and the present procedures that could require modification of the three mechanisms used here. However, if we were to try to adapt and test the present model to lag effects of any kind, initially the same modification would be made to all three mechanisms. Here, we reduced encoding deficit, retrieval, and time in working memory to dichotomous variables that switched, if they switched in a particular procedure, between lag 0 and lag 1. For the lag effect they need to be considered as continuous variables. As the range of lags increase and as there are repeated attempts at retrieval, as there are in the long-term spacing effect, the range of encoding deficit, retrieval, and time in working memory beyond lag 1 can increase dramatically compared with that produced here. This is not to say that the proposed model could be made complete for the long-term spacing effect with this one change. Mechanisms, such as fatigue and perhaps encoding variability, that are not factors in the procedures tested here would probably need to be added in the more complex situation of the long-term spacing effect.

Although we cannot make claims about long real-world lags by studying short laboratory ones, lags of the order used here are of importance outside laboratory tasks. In oral traditions, poetry, and songs, rhyme and alliteration within lines often follow with zero lag to make them more noticeable and to provide immediately sequential cueing of the second member of the pair. Both rhyming and alliterating words between lines are spaced, with the rhythmic pattern making the sound repetition noticeable (Hyman & Rubin, 1990; Rubin, 1995). Hearing the second occurrence of rhyme or alliteration requires the retrieval or continued maintenance in working memory of at least the shared sound pattern from the first. Our matching-first-three-letter task provides an analogue of a process that seems to aid memory outside the laboratory.

Manuscript received 20 May 1996

Manuscript accepted 23 September 1996

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: Advances in research and theory*. (Vol. 1, pp.90-195), New York: Academic Press.
- Baddeley, A.D. (1986). *Working memory*. Oxford: Oxford University Press.
- Bairick, H.P. (1979). Maintenance of knowledge: Questions about memory we forgot to ask. *Journal of Experimental Psychology: General*, 108, 296-308.
- Bairick, H.P., & Hall, L.K. (1991). Lifetime maintenance of high school mathematics content. *Journal of Experimental Psychology: General*, 120, 20-33.
- Bairick, H.P., & Phelps, E. (1987). Retention of Spanish vocabulary over 8 years. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 13, 344-349.
- Bjork, R.A., & Bjork, E.L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.), *From learning processes to cognitive processes:*

- Essays in honor of William K. Estes* (Vol. 1, pp.35–67). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.
- Challis, B.H. (1993). Spacing effects on cued-memory tests depend on level of processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 389–396.
- Cicccone, D.S., & Brelsford, J.W. (1976). Spacing repetitions in paired-associate learning. Experimenter versus subject control. *Journal of Experimental Psychology: Human Learning and Memory*, *2*, 446–455.
- 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., & Watkins, M.J. (1973). The role of rehearsal in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *12*, 599–607.
- Crowder, R.G. (1976). *Principles of learning and memory*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.
- Cuddy, L.J., & Jacoby, L.L. (1982). When forgetting helps memory. An analysis of repetition effects. *Journal of Verbal Learning and Verbal Behavior*, *21*, 451–467.
- D'Agostino, P.R., & De Remer, P. (1973). Repetition effects as a function of rehearsal and encoding variability. *Journal of Verbal Learning and Verbal Behavior*, *12*, 108–113.
- Dellarosa, D., & Bourne, L. Jr. (1985). Surface form and the spacing effect. *Memory and Cognition*, *13*, 529–537.
- Dempster, F.N. (1988). The spacing effect: A case study in the failure to apply the results of the psychological research. *American Psychologist*, *43*, 627–634.
- Ebbinghaus, H. (1964). *Memory: A contribution to experimental psychology*. New York: Dover. [Originally published in 1885.]
- Glanzer, M., & Duarte, A. (1971). repetition between and within languages in free recall. *Journal of Verbal Learning and Verbal Behavior*, *10*, 625–630.
- Glenberg, A.M. (1979). Component levels theory of the effects of spacing of repetitions on recall and recognition. *Memory and Cognition*, *7*, 95–112.
- Glenberg, A., & Adams, F. (1978). Type I rehearsal and recognition. *Journal of Verbal Learning and Verbal Behavior*, *17*, 455–463.
- Glenberg, A.M., & Lehmann, T.S. (1980). Spacing repetitions over 1 8 week. *Memory and Cognition*, *8*, 528–538.
- Glenberg, A., Smith, S.M., & Green, C. (1977). Type I rehearsal: Maintenance and more. *Journal of Verbal Learning and Verbal Behavior*, *16*, 339–352.
- Greene, R.L. (1989). Spacing effects in memory: Evidence for a two-process account. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *15*, 371–377.
- Hintzman, D.L. (1974). Theoretical implications of the spacing effect. In R.L. Solso (Ed.) *Theories in cognitive psychology: The Loyola Symposium*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.
- Hintzman, D.L., Block, R.A., & Summers, J.J. (1973). Modality tags and memory for repetitions: Locus of the spacing effect. *Journal of Verbal Learning and Verbal Behavior*, *12*, 229–238.
- Hintzman, D.L., Summers, J.J., & Block, R.A. (1975). Spacing judgments as an index of study-phase retrieval. *Journal of Experimental Psychology: Human Learning and Memory*, *104*, 31–40.
- Hyman, I.E. Jr., & Rubin, D.C. (1990). Memorabilia: A naturalistic study of long-term memory. *Memory and Cognition*, *18*, 205–214.
- Jacoby, L.L. (1974). The role of mental contiguity in memory: Registration and retrieval effects. *Journal of Verbal Learning and Verbal Behavior*, *13*, 483–496.
- Jacoby, L.L., & Bartz, W.H. (1972). Rehearsal and transfer to LTM. *Journal of Verbal Learning and Verbal Behavior*, *11*, 561–565.
- Kintsch, W., & van Dijk, T.A. (1978). Toward a model of text comprehension. *Psychological Review*, *85*, 363–394.
- Kucera, H., & Francis, W.N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

- Landauer, T.K. (1969). Reinforcement as consolidation. *Psychological Review*, 76, 81–96.
- Landauer, T.K., & Ainslie, K. (1975). Exams and use as preservatives of course-acquired knowledge. *Journal of Educational Research*, 69, 99–104.
- Landauer, T.K., & Bjork, R.A. (1978). Optimum rehearsal pattern and name learning. In M.M. Gruneberg, P.E. Morris, R.N. Sykes (Eds.), *Practical aspects of memory. International conference on practical aspects of memory, Cardiff, Wales*. London: Academic Press.
- Landauer, T.K., & Ross, B.H. (1977). Can simple instructions to use special practice improve ability to remember a fact? An experimental test using telephone numbers. *Bulletin of the Psychonomic Society*, 10, 215–218.
- Modigliani, V., & Hedges, D.G. (1987). Distributed rehearsals and the primacy effect in single-trial free recall. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 13, 426–436.
- Ross, B.H., & Landauer, T.K. (1978). Memory for at least one of two items: Test and failure of several theories of spacing effects. *Journal of Verbal Learning and Verbal Behavior*, 17, 669–680.
- Rubin, D.C. (1995). *Memory in oral traditions: The cognitive psychology of epic ballads, and counting-out rhymes*. New York: Oxford University Press.
- Rubin, D.C., & Friendly, M. (1986). Predicting which words get recalled: Measures of free recall, availability, goodness, emotionality, and pronounceability for 925 nouns. *Memory and Cognition*, 14, 79–94.
- Rubin, D.C., & Wallace, W.T. (1989). Rhyme and reason: Analyses of dual cues. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 698–709.
- Rubin, D.C., & Wenzel, A.E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review*, 103, 734–760.
- Rundus, D. (1971). Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology*, 89, 63–77.
- Thios, S., & D'Agostino, P.R. (1976). Effects of repetition as a function of study-phase retrieval. *Journal of Verbal Learning and Verbal Behavior*, 15, 529–537.
- Toppino, T., & Gracen, T. (1985). The lag effect and differential organization theory: Nine failures to replicate. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 11, 185–191.
- Underwood, B.J. (1970). A breakdown of the total-time law in free recall learning. *Journal of Verbal Learning and Verbal Behavior*, 9, 573–580.
- Watkins, M.J., & Kerkar, S.P. (1985). Recall of a twice-presented item without recall of either presentation: Generic memory for events. *Journal of Memory and Language*, 24, 666–678.
- Waugh, N.C. (1970). On the effective duration of a repeated word. *Journal of Verbal Learning and Verbal Behavior*, 9, 587–595.
- Young, D.R., & Bellezza, F.S. (1982). Encoding variability, memory organization, and the repetition effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8, 545–559.
- Zechmeister, E., & Shaughnessy, J.J. (1980). When you know that you know and when you think that you know but you don't. *Bulletin of the Psychonomic Society*, 15, 41–44.
- Zimmerman, J. (1975). Free recall after self-paced study: A test of the attention explanation of the spacing effect. *American Journal of Psychology*, 88, 277–291.

