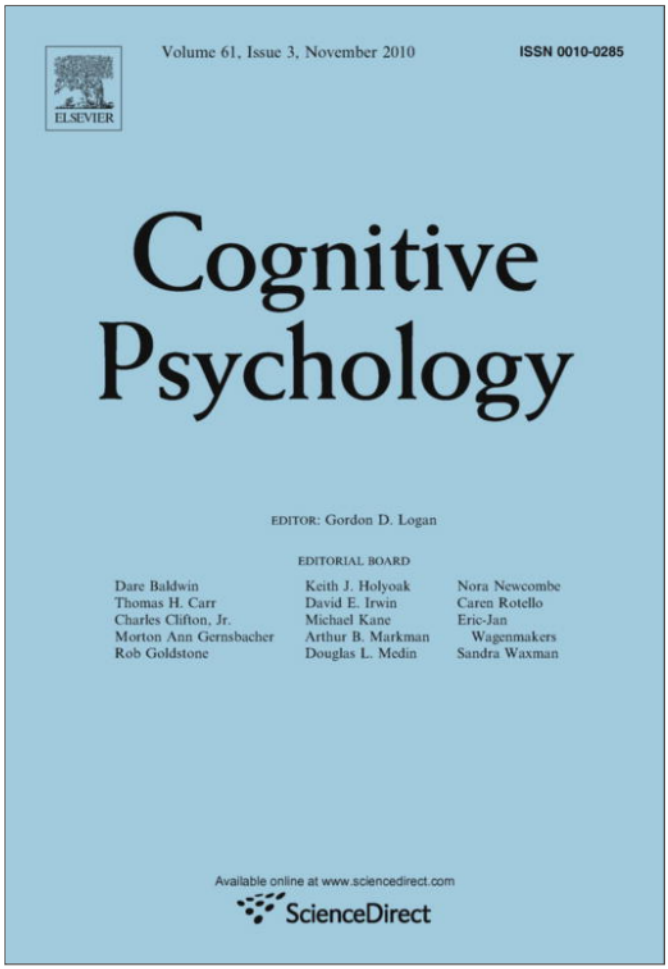
Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright

Author's personal copy

Cognitive Psychology 61 (2010) 228–247

Contents lists available at ScienceDirect

Cognitive Psychology

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

What makes distributed practice effective?

Aaron S. Benjamin *, Jonathan Tullis

Department of Psychology and the Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, United States

ARTICLE INFO

Article history:
Accepted 24 May 2010

Keywords:
Spacing effects
Lag effects
Explicit memory
Encoding variability
Reminding

ABSTRACT

The advantages provided to memory by the distribution of multiple practice or study opportunities are among the most powerful effects in memory research. In this paper, we critically review the class of theories that presume contextual or encoding variability as the sole basis for the advantages of distributed practice, and recommend an alternative approach based on the idea that some study events remind learners of other study events. Encoding variability theory encounters serious challenges in two important phenomena that we review here: superadditivity and nonmonotonicity. The bottleneck in such theories lies in the assumption that mnemonic benefits arise from the increasing independence, rather than interdependence, of study opportunities. The reminding model accounts for many basic results in the literature on distributed practice, readily handles data that are problematic for encoding variability theories, including superadditivity and nonmonotonicity, and provides a unified theoretical framework for understanding the effects of repetition and the effects of associative relationships on memory.

© 2010 Elsevier Inc. All rights reserved.

1. Introduction

Recent years have seen a resurgence of interest in the advantages provided to memory by the distribution of multiple practice or study events. This interest owes in part to the ramifications such results have for developing effective training, educational, and athletic regimens, and to the impressive success researchers have had in importing interventions out of the laboratory and into training settings, the classroom, and the practice field. The effects of distributing practice are extremely robust and cross-cutting—the advantages are evident in basic memory tasks using words (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Janiszewski, Noel, & Sawyer, 2003) and pictures (Hintzman & Rogers,

* Corresponding author.
E-mail address: ASBENJAM@ILLINOIS.EDU (A.S. Benjamin).

0010-0285/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.cogpsych.2010.05.004

Author's personal copy

1973), in motor skill acquisition (Lee & Genovese, 1988), and with more complex educationally relevant materials (Krug, Davis, & Glover, 1990). Almost every major textbook on cognitive psychology and on human memory discusses the advantages of distributed learning, and such effects are among the keystone results that contending theoretical models of learning and memory must dutifully explain.

The results from experiments on distributed practice have informed memory theory principally insofar as those theories successfully handle the difference between massed and spaced practice conditions (the spacing effect; Ebbinghaus, 1964) or between differentially spaced conditions (the lag effect; Melton, 1967), as well as how they account for differences arising from intentionality of learning during study or from different memory tests (Challis, 1993; Greene, 1989; Russo, Parkin, Taylor, & Wilks, 1998). Here we make the case that the superadditivity effect (Begg & Green, 1988; Glanzer, 1969; Watkins & Kerkar, 1985; Waugh, 1963), which concerns the relationship between memory for single events and memory for multiple events, and the nonmonotonicity effect, which concerns the shape of the lag function, are empirical hallmarks that are equally important but less widely acknowledged, and that the most prominent class of theoretical explanations of distributed practice faces challenges in accounting for either result. On the other hand, a theory based on the concept of reminding (Benjamin & Ross, in press; Hintzman, 2004) provides a means of handling these effects quite naturally.

In this paper, we first review encoding variability accounts of distributed practice and contrast that perspective with other extant theories. We argue that the standard-bearer for such theories, Stimulus Sampling Theory (SST; Estes, 1955a, 1955b) makes a strong prediction about the relationship between memory for once-studied and memory for twice-studied materials and then provide a simple mathematical proof of that claim. The relationship of the independence baseline to memory performance under repeated conditions has been investigated elsewhere (Begg & Green, 1988; Watkins & Kerkar, 1985; Waugh, 1963), but the relevance of such data to the quite widespread acceptance of encoding variability as a basis for distributed-practice effects has not been previously considered.

We then relax some of the assumptions of Stimulus Sampling Theory and consider more generally the shape of the lag function that encoding variability theories can account for. By extending the machinery of SST to cases in which retention interval is varied, we demonstrate that encoding variability theory—even with very liberal assumptions about response mechanisms—requires substantial modification from SST before it can account for a nonmonotonic relationship between lag and performance.

In the second portion of the paper, we review data relevant to the presence of superadditivity and nonmonotonicity in lag functions. We report one meta-analysis of the relevant literature that reveals superadditivity to be ubiquitous in extant research on spacing. We also review specific experiments from the literature that effectively confront the many potentially confounding variables in repetition paradigms, and conclude that superadditivity and nonmonotonicity are reliable, replicable, and important features of spacing functions. Throughout, we review the ways in which encoding variability theory must depart from the assumptions of SST in order to accommodate these results.

Finally, in the third section, we consider a simple, alternative theoretical perspective based on the concept of reminding. Such an approach produces superadditive, nonmonotonic functions quite readily, and also accounts for other major results in the literature. We propose a simple statistical model that features the qualities a model of reminding needs to be viable, and consider the extensive implications for how we think about the beneficial effects of distributing practice, as well as how prominent extant models of memory (e.g., Raaijmakers & Shiffrin, 1980; Shiffrin & Steyvers, 1998) should be generalized to accommodate the effects of repeated presentations on memory.

2. Contextual and encoding variability

The basic principle underlying encoding variability theory is that aspects of the encoding process or contextual circumstances vary with time, and that events that are further apart in time are thus likely to be more "different" than events that are close together. The variability may refer to contextual elements that become associated with the to-be-remembered stimulus or to the encoding processes that

Author's personal copy

are applied to the stimulus and representations that are yielded as a consequence (Bower, 1972; Estes, 1955b), or even to the relative proximity of the underlying traces in a searchable memory space (Landauer, 1975). Memory benefits in any of these scenarios because the greater variability confers a greater likelihood of generalization to the circumstances of testing. This may occur because of more routes to effective retrieval, or because aspects of those representations are more likely to overlap with contextual or mnemonic elements at test.

Historically, such theories originated with the Stimulus Sampling Theory proposed by Estes (1955a, 1955b). That theory was originally applied to problems in acquisition and extinction in animal learning, but variations on it were generalized to distributed practice in human verbal learning by Bower (1972), Madigan (1969), and Melton (1970). This theoretical perspective is endorsed in major textbooks (Haberlandt, 1999; Sternberg, 2003) and in historical (Crowder, 1976) and current (Cepeda et al., 2006) reviews of the literature. It is quite likely the predominant current explanation for spacing effects.

Such theories imply that increasing performance as a function of spacing derives from an increasing degree of independence between the individual study events. SST thus predicts monotonically increasing levels of performance with spacing that asymptote at a level that is consistent with an assumption of additivity of the individual probabilities (henceforth, the independence baseline), as shown in the next section. These two qualities—monotonicity and additivity—are implications of the presumption that performance benefits from decreasing dependence between the study events, and that those events can be correlated no less than 0.[1] Some authors have noted levels of performance that are considerably higher than the independence baseline (Begg & Green, 1988; Crowder, 1976; Watkins & Kerkar, 1985; Waugh, 1963), but others have claimed that performance does likely asymptote at that baseline (Glanzer, 1969; Paivio, 1974). One purpose of the forthcoming analysis is to evaluate this question meta-analytically.

3. Stimulus Sampling Theory and performance additivity

When two practice or study events are independent of one another, performance or memory for that event should reflect a statistical relationship of independence between those events. Given the probability $p_{(1)}$ of recalling a singly studied item from a list, the independence rule predicts levels of recall for two presentations ($p_{(2)}$) as 1 minus the probability of not recalling either presentation:

$$p_{(2)} = 1 - (1 - p_{(1)})^2.$$

This simple relationship has been noted in the context of spacing effects by numerous authors (Crowder, 1976; Watkins & Kerkar, 1985; Waugh, 1963). Mostly, it has been discussed because it often underpredicts performance, as described in greater detail below.

Stimulus Sampling Theory (Estes, 1955b) describes performance as a function of the proportion of contextual elements encoded during each of m study exposures, the interval between those exposures, and the rate with which those elements fluctuate in availability:

$$F(m) = 1 - (1 - J)(1 - F[0])(1 - J + Ja^t)^{m-1}, \quad (1)$$

in which J is the proportion of elements available to be encoded that are included in any single encoding event, $F[0]$ is the probability of remembering the event given no study exposure, a represents the rate of fluctuation among the contextual elements, and t is the duration of the interstudy interval.

When there is only a single study event ($m = 1$) and the probability of remembering an unstudied item is (quite reasonably) assumed to be 0, then

$$F(1) = J,$$

[1] Strictly speaking, this is only true when the underlying contextual elements are themselves independent of one another, as they are in the theories that have been applied to spacing. If those elements are reconceptualized as correlated oscillators, then the correlation across time period can fluctuate more dramatically and drop below 0 (and thus violate the conditions leading to the independence baseline). However, any amount of noise in such oscillators will still lead the correlations to drop in magnitude and eventually asymptote at 0.

Author's personal copy

and the predicted level of performance for a twice-studied item under the assumption of independence is[2]

$$\text{predicted } F(2) = 1 - (1 - J)^2.$$

This equation represents the application of an independence rule to the predicted level of performance given a single presentation. It can now easily be shown that the model predicts this same level of performance asymptotically when two study events are spaced. Performance on a twice-studied item is a function of the proportion of elements available during each exposure, the rate of fluctuation, and the interval between the two study events:

$$F(2) = 1 - (1 - J)(1 - F[0])(1 - J + Ja^t). \tag{2}$$

If $F(0)$ is assumed as before to be 0 and $-1 < a < 1$ (that is, if fluctuation in and out of the current context is presumed to neither be nonexistent nor complete, neither of which is plausible, nor would yield any kind of spacing effect), then

$$\text{as } t \to \infty,\ Ja^t \to 0,$$

and the previous equation reduces to

$$F(2) = 1 - (1 - J)^2,$$

exactly as predicted by the independence rule.

Thus, SST is incompatible with superadditive lag functions. However, superadditivity does not deny the viability of other possible variants of encoding variability theory. In particular, this demonstration is only valid under the particular response assumptions made by Stimulus Sampling Theory—in particular, the assumption that the probability of a positive response is linearly related to the amount of overlap between encoded and available contextual elements during the test. If we changed the nature of the response variable such that a positive response was, for example, a step function of $F(m)$, then the mathematics of the proof no longer generalize to the response variable. In the next section, we evaluate the extant literature meta-analytically and conclude that superadditivity is ubiquitous and, consequently, that SST does not provide a correct account of the beneficial effects of distributed practice. We then consider how the response functions of SST can be generalized in such a way so as to handle superadditivity, and what inferences can be drawn from such generalized models.

4. Data collection

Potential articles were gathered from three separate sources. First, a list of potential articles was compiled by searching PsycINFO (1872–2008) for the following relevant keywords: "spacing effect," "distributed practice," "spac* mass* practice," "spac* mass* learning," "spac* mass* presentation," "spac* mass* retention," "mass* distrib* retention," "spac* remem*," "distrib* remem*," "lag effect," "distrib* lag," and "distrib* rehears*." Second, the reference list from a recent meta-analytic review of the spacing effect (Cepeda et al., 2006) was examined for potential studies. Finally, the citation lists of all of the previously gathered articles were examined to identify any potential articles that may have been missed by the other search methods.

Studies were included only if they met two major criteria. First, the studies had to use either written, verbal, or pictorial materials (paired associate terms, lists of words, texts and paragraphs, sentences, pictures, etc.) and explicit memory tests (either recall or recognition). Studies were included only if memory performance was reported or could be computed from the data provided. Implicit memory tests and tests involving skill acquisition were excluded. Second, experiments must have reported performance following both single learning trials and double learning trials, so that single and double presentation performance rates could be compared. Studies reporting only one of these two

[2] This assumption may be incorrect in experimental designs in which list position is not effectively controlled over p_1 and p_2. In such cases, the effect of a difference between p_1 and p_2 can underestimate the independence baseline, but the difference between p_1 and p_2 must be very large to effect any noticable change to that estimate.

Author's personal copy

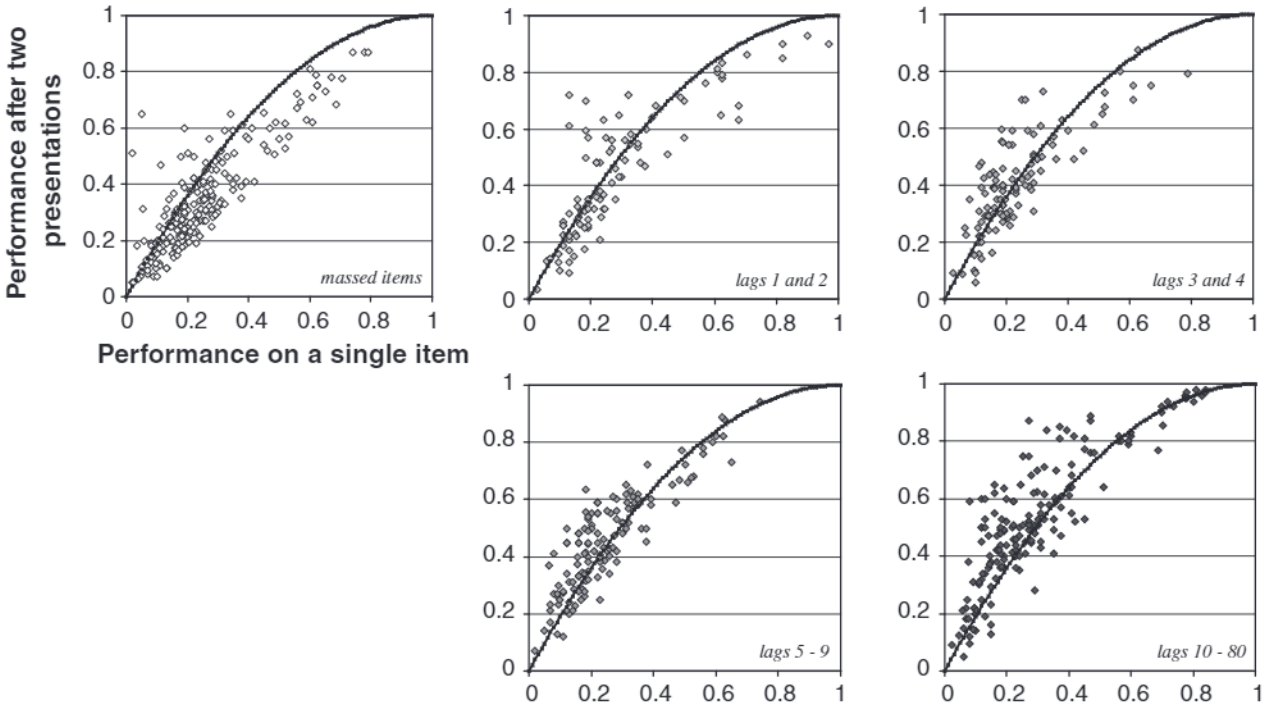


Fig. 1. Actual (diamonds) and predicted (black line) performance for repeated items as a function of performance on items studied once for all conditions in the data set for which lag was precisely specified. Longer lags are indicated by darker shading of the diamonds.

values were necessarily excluded from analysis. Experiments using children, adults, and clinical populations were included. Studies were not excluded on the basis of intentionality at the time of study; intentional and incidental learning paradigms were both included. The entire dataset consists of 829 different conditions from a total of 72 articles. A list of the articles, experiment numbers, and conditions included in the analysis is available at: http://www.psych.illinois.edu/≃asbenjam/SuperadditivityReferences.htm.

5. Analysis of spaced practice and independence baseline

First, we evaluate performance on twice-studied items as a function of lag relative to the independence baseline. Fig. 1 shows the data from a subset of 735 different conditions, grouped into five lag ranges. This analysis includes all of the conditions for which the specified lag or lag range fit within the lag intervals chosen in terms of number of intervening items (including massed items, which are indicated as lag 0).[3]

It does not include conditions in which the lag was varied over a range of values that spanned two of our specified lag ranges, nor conditions for which the lag was not specified. Table 1 indicates how many data points contributed to each level of spacing evaluated in this analysis and the proportions of items within each lag range that exceed the independence baseline. Increasingly spaced repetitions are shown with increasingly dark shading in Fig. 1. The proportion of data points that lie above that line increases with lag, and lags of greater than five items induce levels of performance that are above the independence baseline at a very high rate, as indicated in Table 1. Similarly, mean levels of actual performance lie above the predicted levels for those latter two lag intervals, as shown in Fig. 2 ($t[134] = 3.76$; $t[171] = 6.36$).

The evidence from this analysis is very clear in revealing that performance does not asymptote at the independence baseline. At lags as short as five intervening items, a majority of data points lie

[3] Naturally, the magnitude of spacing effects does not vary simply with lag, but rather as a function of the relationship between lag and retention interval (Cepeda et al., 2006; Glenberg, 1976). However, there is too little variability in retention intervals across studies to evaluate superadditivity as a joint function of these variables meta-analytically.

Author's personal copy

Table 1
Number of data points contributing to each level of spacing and the proportion of data points that numerically exceed the independence baseline.

Range	Count	Proportion exceeding independence baseline
All test types		
Massed	218	.17
1–2	98	.31
3–4	112	.46
5–9	135	.61*
10–80	172	.69*
Cued recall only		
Massed	31	.16
1–2	12	.58
3–4	18	.28
5–9	11	.18
10–80	66	.62*

* Indicates that a probability of this magnitude of larger is $p < .05$ as assessed by the binomial distribution.

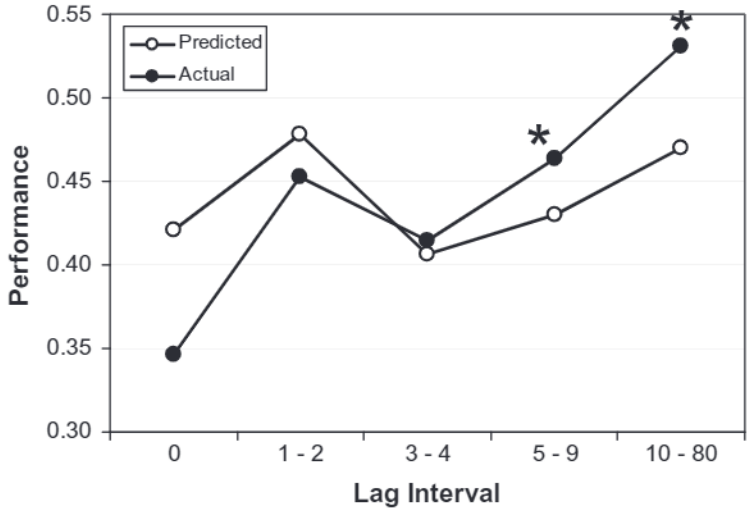


Fig. 2. Mean actual (black circles) and mean predicted (open circles) performance as a function of the lag bands used in this figure. Asterisks indicate that performance is higher than predicted by the independence baseline ($p < .001$).

above that baseline, and mean performance is well above the predicted mean level. However, one serious concern in such an analysis is that of a list-strength artifact (Ratcliff, Shiffrin, & Clark, 1990). Because the vast majority of experimental conditions involve tests of free recall (583 cases), it is quite possible that performance in the single-item condition (and estimates of the independence baseline) is artifactually depressed by competition with stronger, repeated items. Because this competition appears to occur during output (Baüml, 1997), tests such as cued recall and recognition do not elicit the list-strength effect and should be free from such an artifact. In this next analysis, we address this concern.

Although there were not a sufficient number of recognition conditions in the database to permit examination, there was a reasonably high number of conditions that utilized cued recall (148 in total), and we examined this subset of cases for evidence of superadditivity. This analysis is shown in Fig. 3, and reveals a similar but somewhat smaller effect. Only items in lags greater than 10 lie above the independence baseline at a statistically significant rate. In addition, the mean level of cued recall performance on items in this long lag range (0.58) is higher than predicted by the independence baseline

Author's personal copy

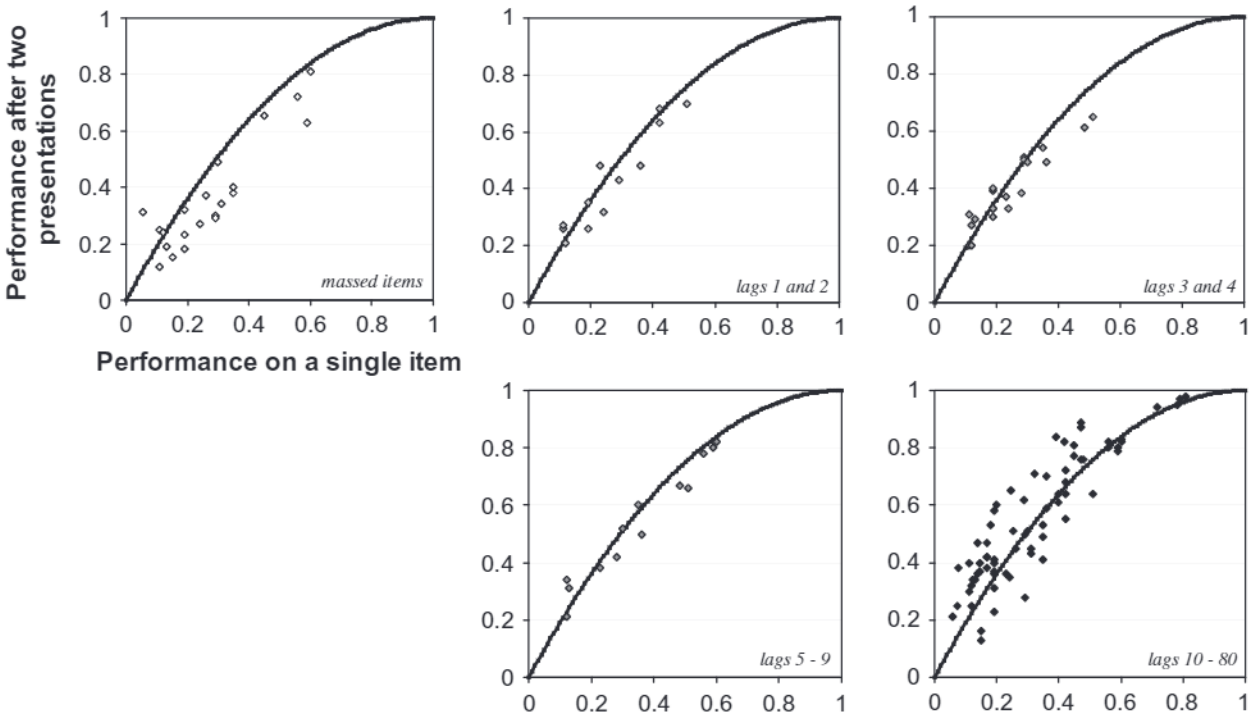


Fig. 3. Actual (diamonds) and predicted (black line) performance on repeated items as a function of performance on items studied once for all conditions in the data set for which a cued-recall test was administered. Longer lags are indicated by darker shading of the diamonds.

(0.54; $t[65] = 3.29$). This result demonstrates that list-strength competition does not account for the rejection of the independence baseline. It should also be remembered that, in a meta-analysis of this type, the artifacts introduced by averaging across subjects (that is, by applying the nonlinear independence rule to group, rather than individual data) necessarily overestimates the independence baseline, and consequently works against the detection of superadditivity (Begg & Green, 1988).

Stimulus Sampling Theory is incompatible with the finding that performance on twice-presented items consistently lies above the level predicted by the independence baseline. The fact that this result obtains in cued as well as free recall suggests that list-strength confounds do not underlie this effect. However, another potential confound in repetition paradigms is the input serial position (or equivalently, recency) of singly presented and twice-presented items. Particularly at long lags, it is very difficult to ensure that the second presentation of a repeated item occurs no later that the average presentation of a singly presented item. This confound is problematic here because a recency confound could artificially inflate performance for repeated items. Because the degree of and methods for control over equating serial position across conditions differ widely across different studies, it is difficult to conditionalize our meta-analysis on studies that have successfully accounted for this confound. However, one very major study (Begg & Green, 1988) sought evidence for superadditivity and implemented exquisite controls for serial position; we review their methodology and results here.

As noted earlier, one potential problem in studies of repetition and spacing is the output interference that can exaggerate the advantage of superior encoding conditions. Our earlier solution was to examine conditions in which output order is controlled, like tests of cued recall or recognition. A second problem is the potential for different mean input serial positions for once-presented and twice-presented items. It can be difficult to control for recency between these conditions, but some studies do so effectively. The first experiment of Begg and Green (1988) provides a particularly good example of an experiment that controls for both of these potential confounds. They had subjects study word pairs and tested them using cued recall. As noted before, the use of cued recall eliminates concerns about output order. Second, they controlled for serial position by implementing between-subject control conditions that placed single presentations in the positions that pairs occupied for other subjects. The one group of subjects that experienced repetition viewed a study list in which singly presented

Author's personal copy

items were evenly distributed throughout the two halves of that list, as were the repetitions. This provides one control: randomization of the list haves ensures that mean serial position for second repetition of an item and for the nonrepeated items presented in the second half of the list is roughly equivalent. They also included two additional between-subject control groups in which no pairs were repeated. In each of those conditions, a new set of singly presented items replaced either the first or the second presentation of the repeated items for the repetition group.

In such a design, there are five ways that the independence baseline can be computed. It can be based on the common nonrepeated items presented in the first and second halves of the list, once for each of the three between-subject conditions. In addition, it can be based on level of performance for the exact items that are repeated for one group of subjects by examining performance on those items for subjects who experienced them only once (in either the first or second list half, yielding two more comparisons). All five of these comparisons revealed superior performance for repeated items than the level predicted by the independence baselines. In fact, because this study included another between-subjects variable not relevant to our present considerations, the design yielded ten possible comparisons, all of which revealed superadditivity. This particularly powerful and well controlled demonstration enhances our confidence that the results of the meta-analysis are not attributable to the inclusion of studies with poor control over input position and output interference.

However, the relevance of this demonstration to encoding variability theory relies on the particular response assumptions made by Stimulus Sampling Theory. We turn next to an analysis of nonmonotonicity, which is not predicted by encoding variability theory under any plausible response assumptions.

6. Stimulus Sampling Theory and nonmonotonicity

Although it is necessary to specify the function relating contextual overlap with response probability to draw predictions about the relationship between memory for singly presented items and memory for repeated items, other aspects of performance can be derived even under quite relaxed response assumptions.

Let us consider the implications of relaxing the assumption of linearity in a way that does not do serious violence to the underlying principles of encoding variability theory. First, increasing overlap of elements should be at least weakly monotonically related to response probability. That is, it should never be the case that a higher amount of overlap between encoded contexts and current test context leads to a lower probability of recall. So, our generalization of Eq. (1) treats $F(m)$ as a latent variable, and R as the response function:

$$R = f[F(m)],$$

where $f(\cdot)$ is a weakly monotonically increasing but otherwise unspecified function.

First, we will consider the implications of such a generalization and then move on to a formal proof of those characteristics. Because increasing the lag between p_1 and p_2 serves only to increase the diversity of encoded elements, $F(m)$ can only increase with lag. Because R is a monotonic function of $F(m)$, therefore, it can never decrease with lag. Thus, the shape of the response function can never be nonmonotonic, even when response thresholds or other highly nonlinear response functions are introduced. This point is critical because one benchmark datum in the spacing literature is the interaction between retention interval and lag demonstrated by Glenberg (1976). His results showed obvious nonmonotonicity at short lags, and he noted that encoding variability theory provided an effective theoretical position to account for such an effect.

To evaluate directly whether SST can account for lag function nonmonotonicity, it must be generalized to the case in which retention intervals are variable. To do so, we assume that elements continue to fluctuate between the second presentation of an item and the test of that item and label that interval RI. Then, using only parameters from the original model (Eqs. (1) and (2)), the generalized function is:

$$R(RI, t, m = 2) = J[J + (1 - J)a^{t+RI}] + (1 - J)[J(1 - a^{t})][J + (1 - J)a^{RI}], \tag{3}$$

a proof of which is provided in Appendix A.

Author's personal copy

To evaluate whether this model is capable of producing nonmonotonic response functions, and thus dealing successfully with the interaction between retention interval and lag, the derivative of the function is taken with respect to t:

$$\frac{d}{dt}R(RI,t,m=2)=J^2a^t\ln a(J-1-Ja^{RI}+a^{RI}), \tag{4}$$

and this function is solved for 0 to estimate values of t where inflection points will occur. Doing so yields a solution only when RI is zero, indicating that nowhere within the meaningful bounds of this function is there an inflection point. This derivation is provided in Appendix B.

This analysis demonstrates that, even under the most liberally relaxed response assumptions, SST cannot accommodate lag functions that are nonmonotonic. Although nonmonotonic lag functions have been reported in the literature and their relevance to encoding variability has been noted (most thoroughly by Crowder (1976)), we sought evidence in the literature for nonmonotonic lag functions under conditions in which, just as before, the potential for confounding of input serial position and output interference are minimized. A recent report by Cepeda, Vul, Rohrer, Wixted, and Pashler (2008) provides an excellent example, and we review the relevant aspects of their work here.

In their very large internet-based study, subjects learned a series of facts and were given an opportunity to re-study those facts at a later date. Then, yet later, their memory was tested via cued recall and forced-choice recognition. Though both of these procedures effectively deal with the problem of output interference, we discuss here only the results from the cued-recall test, since it was administered prior to the recognition test (which included the same items). Critically, and much unlike the modal paradigm in which spacing is studied, the retention interval between the re-study occasion and the test was controlled independently of lag between the two study sessions; that is, the interval between the second study occasion and the test was controlled.

There are many important and commendable aspects of the Cepeda et al. (2008) study, but the important finding for present purposes is that they showed very clear evidence for nonmonotonic lag functions. They also replicated the important findings of prior researchers (Glenberg, 1976; Peterson, Hillner, & Saltzman, 1962) that the optimal interval between study events for retention increased with the interval prior to the test. Taken together, these three studies demonstrate quite clearly that nonmonotonic lag functions can be quite easily found, and that there is a systematicity to the form of those functions that any respectable theory of spacing should account for.

7. Theoretical approaches to understanding distributed practice

Current theories of the effects of spacing vary in their similarity to Stimulus Sampling Theory, and it is fair to ask what bearing the findings of superadditivity and nonmonotonicity have on the viability of those approaches. We consider here some of the more prominent classes of alternatives, but it should be kept in mind that the majority of these theoretical suggestions have not been computationally implemented and are consequently open to interpretational differences. These theoretical alternatives are considered more broadly by Cepeda et al. (2006); we focus here specifically on the ability of such theories to accommodate nonmonotonic functions and superadditive levels of performance.

7.1. Encoding variability

As we have indicated throughout, although SST is incapable of producing nonmonotonic, superadditive lag functions, there are versions, or potential versions, of encoding variability theory that are not bound to those constraints. We already considered how relaxing the linear response mechanism can allow encoding variability theory to yield superadditive performance; encoding variability theory can also produce nonmonotonic lag functions when very different assumptions are made about the nature of the encoding functions than those employed by SST.

In SST, any contextual attribute that is experienced during either presentation event is fully encoded. We shall call this assumption the *all-or-none element encoding assumption*, indicating that elements are either un-encoded or fully encoded. If an element is encoded during one of those events,

Author's personal copy

it does not matter if it is encoded again during a later event—performance does not change. An implication of this characteristic is that, at short lags, the high redundancy between elements available during the two presentations underlies the lower levels of performance. But, since the overlap between elements at test and elements at p_1 cannot be regularly greater than the overlap between elements at test and elements at p_2, performance will always be highest at the longest lags.

If one abandons the all-or-none element encoding assumption, either by assuming continuous learning functions for contextual elements, or by adopting continuous-valued rather than Bernoulli elements, then the redundancy problem is mitigated and nonmonotonic functions are possible. Examples of such models include the one outlined by Glenberg (1976) and the recent ones provided by Mozer, Pashler, Cepeda, Lindsey, and Vul (2009) and Raaijmakers (2003), though neither of the latter are pure encoding-variability models.

It seems that any theory in which the lag between repeated presentations increases the dissimilarity between representations of those events, and in which dissimilarity benefits performance—encoding variability theories, in short—must possess two characteristics that distinguish it from SST if they are to be able to deal with superadditivity and nonmonotonicity. First, they cannot include the all-or-none element encoding assumption. Second, the response function must be a nonlinear function of the amount of overlap between encoded elements and elements available at test.

Conceptually, such theories presume that greater generalization results from the greater diversity of representations accorded to repeated items spaced further apart in time. We do not dispute this point, but do emphasize that the findings of superadditivity and nonmonotonicity suggest that more is needed to understand the effects of spaced repetitions on memory.

7.2. Diminished processing

Another class of theories attributes the beneficial effects of spacing to the attenuation of normal encoding processes when lags are short. The inattention theory of Hintzman (1974) exemplifies this view. Diminished processing explains why short lags suffer relative to long ones, but has little to say about the determinants of performance at long lags. However, tacit in the idea of diminished processing is the idea that the processing of the second presentation would be "full"—i.e., identical in scope to the first—were it not diminished by familiarity or inattention. In other words, this theoretical perspective seems to imply that eventual performance will be best when processing of the two events is independent of one another. The implication that independent encoding of the two presentations is the optimal case leads to the same difficulties with superadditive performance as discussed earlier.

Unless diminished processing is also thought to occur at very long intervals, it is also not obvious how such theories could deal with nonmonotonic lag functions. Thus, diminished processing, on its own, would seem to be unable to deal with either major phenomenon discussed here.

7.3. Study-phase retrieval

A third suggestion is that the benefits of spacing arise from the effects of retrieving elements of the first presentation at the time of the second (e.g., Hintzman, 2004; Hintzman, 2010; Thios & D'Agostino, 1976). Because this suggestion places the interaction between, rather than the independence of, the two presentations at the heart of distributed-practice effects, it has the potential to avoid the pitfalls of asymptotically independent levels of performance. However, it would seem that any theory in which study-phase retrieval is the only mechanism by which repetition improves memory is stymied by the direction of the spacing effect: because retrieval is more likely at short lags, the beneficial consequences of that retrieval should be greater at short lags. This prediction is obviously at odds with the empirical regularity of the effect.

Some theories evade this trap by supplementing study-phase retrieval with additional mechanisms. The theory of Greene (1989) augments a study-phase retrieval mechanism with both contextual variability and deficient processing under massed conditions. The relevant aspect of the theory for the current analysis is that study-phase retrieval engenders encoding of the contextual elements that have changed between the two presentations. More such elements are stored when the lag is greater, and later retrieval of the contents of memory is more likely when more elements are stored.

Author's personal copy

Because the encoding of contextual elements in this theory provides information above and beyond what two independent encodings would provide, it could conceivably account for superadditive performance. However, as we demonstrate below, the same effect can be achieved simply by considering the mnemonic consequences of retrieval. According study-phase retrieval theory with such mechanisms also provides for nonmonotonic lag functions, as will be evident below. Thus, one can eschew entirely the aspects of the theory that appeal to encoding variability and achieve the same result somewhat more simply.

7.4. Consolidation

Wickelgren (1972) presented a general theory of long-term memory in which the resistance of memories—a factor that offsets the effects of forgetting—was thought to be greater for the second of two presentations of a stimulus. Spacing was predicted to improve memory because the second presentation inherited the degree of resistance achieved at that point by the first trace, and that amount was presumed to increase with time. It would not seem that consolidation theory can deal with the nonmonotonic patterns evident in the interaction between retention interval and lag, but it can naturally predict superadditivity because the retarded decay rate will lead to superior retention relative to two presentations with equal decay rates (as might be presumed for an independent encoding). As formulated, however, Wickelgren's (1972) theory encounters difficulty handling other phenomena in the literature on distributed practice (e.g., Hintzman, Block, & Summers, 1973). Nonetheless, the general prediction that repeated presentations has the potential to affect the functional rate of forgetting is one that has been considered more generally (e.g., Bjork & Bjork, 1992; Pavlik & Anderson, 2005), and we will revisit it when considering the interaction between retention interval and spacing.

8. General principles of a model of distributed practice

The fact that spacing yields asymptotically superadditive levels of performance suggests that a viable theory would start with the following three characteristics. First, the processing that occurs on the two study occasions must be of an interactive, rather than an independent, nature. Second, that interactivity must be of a positive sort, in order to support superadditive, and not subadditive, performance. Third, that interactivity must support increasing and then decreasing benefits with spacing.

Theories that postulate contextual or encoding variability as a basis for the benefits of distributed practice encounter some difficulty on the first benchmark, as shown here. SST attributes superior performance to increasing independence of the practice events and requires modification to account for superadditive performance and nonmonotonic lag functions. The class of theories that postulates deficient processing under massed conditions is at the very least incomplete, in that it proposes a basis for performance lower than the independence baseline (the deficient processing induced by short lags) but contains no mechanism that explains why spaced presentations lead to superadditive levels of performance. That is, the fact that the interactivity between presentations only supports subadditivity indicates that they fail on the second benchmark.

Theories that involve study-phase retrieval (Hintzman, 2004; Hintzman, 2010; Pavlik & Anderson, 2005; Thios & D'Agostino, 1976) implicate the retrieval of the first event during the second presentation as a basis for superadditive levels of memory. If it is assumed that that retrieval benefits memory to the degree that it is difficult (Bjork, 1988; Pavlik & Anderson, 2005), or that deficient processing is induced at short intervals (Greene, 1989), then such theories also meet the third benchmark. The combined effects of study-phase retrieval and retrieval difficulty explain why manipulations that introduce interference between the study events enhance performance (Bjork & Allen, 1970) and why memory for the contextual elements of the first presentation increase with short lags and decrease at later lags (Appleton-Knapp, Bjork, & Wickens, 2005).

9. A reminding-based statistical model of distributed practice

In this section, we outline a general framework for how the principles espoused here as central to any successful model of distributed practice could be incorporated into a simple model of

Author's personal copy

performance. We do not provide here a serious process model of the effects of repetition; rather, we outline a simple statistical case that highlights the ways in which process models, such as Minerva2 (Hintzman, 1988) or SAM (Raaijmakers & Shiffrin, 1981), could be augmented to treat the beneficial effects of repetition as a result of reminding. These principles supplement the basic tenets of study-phase retrieval theory (e.g., Greene, 1989; Hintzman, 2010; Thios & D'Agostino, 1976) with simple, generally accepted concepts from the memory literature. There are three basic principles that underlie the *reminding model* we instantiate here:

(1) Items presented at t_1 are subject to forgetting with time ($F[t]$). For this, we postulate power-law forgetting (Anderson & Schooler, 1991; Wixted & Ebbeson, 1991).
(2) Items presented at t_2 vary in their capacity to spontaneously elicit reminding of the event at t_1 ($R[p_1|p_2]$). Repetitions are presumed to enjoy high capacity for reminding, associates somewhat less, and unrelated items very little.
(3) The act of retrieval potentiates memory, and the degree to which it does so is positively related to the difficulty of the retrieval. That is, successful reminding following either a high degree of forgetting or a low amount of reminding potential enhances memory more for the retrieved information than would a reminding following little forgetting or high reminding potential.

This view specifies the probability of reminding as:

$$RP = R(p_1|p_2)F(t),$$

where

$$F(t) = t^{-k},$$

t is the positive interval ($t \geqslant 1$) between p_1 and p_2 and k is the positive rate of forgetting. The potentiating value of a successful reminder is

$$1 - RP.$$

The net effect of these offsetting factors is set to be a constant proportion of the level of memory for a singly presented item, thus assuring that performance will lie between 0 and 1.

$$M(p_1) = p_1 + (1 - p_1)RP(1 - RP). \tag{5}$$

where M represents the final level of memory performance for the presentation at p_1 after an additional presentation at p_2 (or the presentation of an associate at p_2).

Under the simplifying assumption that $R(p_1|p_2) = 1$ when the item is simply repeated, the model for repeated presentations is:

$$M(p_1) = p_1 + (1 - p_1)t^{-k}(1 - t^{-k}),$$

and contains only one free parameter (k).

There are numerous ways in which this set of assumptions considerably oversimplifies the memory task in a distributed practice paradigm (for example, by assuming no forgetting following t_2), but it will serve adequately to illustrate how a very simple and mostly uncontentious set of principles can easily handle basic findings that have proven difficult for theories based on encoding variability. To apply to any real set of data, many aspects of the model, such as the weighting of the tradeoff between reminding probability and reminding potential, would need to be more fully parameterized. The model presented by Pavlik and Anderson (2005) provides a fuller realization of how reminding can be implemented in a process model.

9.1. Superadditivity

First, we can demonstrate that this model can (but does not always) predict superadditivity following repeated presentations. Performance of the model under multiple conditions relative to the independence baseline is shown in Fig. 4. There it can be seen that actual memory for twice-presented

Author's personal copy

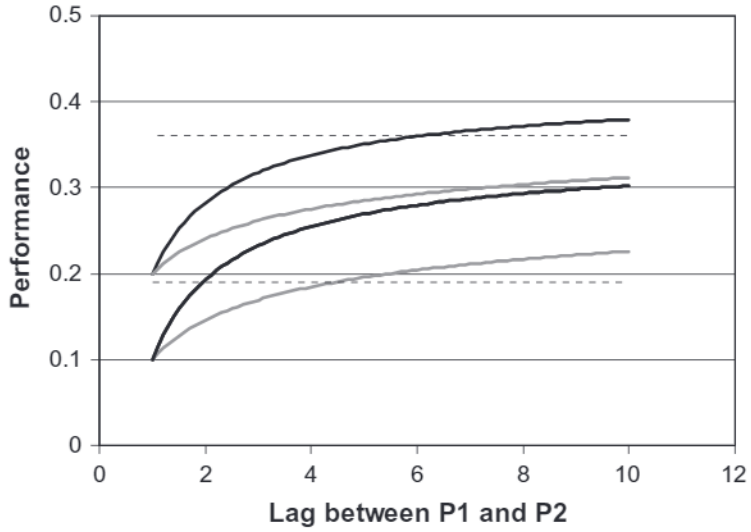


Fig. 4. Demonstration of superadditivity by the reminding model. Solid lines indicate performance on twice-presented items, with darker lines indicating faster forgetting. The dashed lines indicate the independence baseline. Two different levels of p_1 performance are displayed.

items is sometimes above and sometimes below the independence baseline (dashed line). This effect varies with single-item performance (higher levels of memory for singly presented items decrease the likelihood of superadditivity) and forgetting rate (the portions of the curve that exceed the independence baseline depend on the rate of forgetting, which is greater for the darker than the lighter curves).

9.2. Nonmonotonicity

Because performance in the reminding model represents a tradeoff between the probability of successful reminding and the value of that reminding, the resultant function will always be nonmonotonic (although the inflection point may occur at some lag outside the range of empirical data for a given experiment). As noted previously, data from spacing experiments often reveal nonmonotonic relationships between lag and performance (e.g., Peterson, Wampler, Kirkpatrick, & Saltzman, 1963), and this is another datum with which encoding variability theory cannot easily be reconciled. Because there is no process in such theories that relates forgetting of the first presentation to the encoding processes implemented at the time of the second presentation (cf. Greene, 1989), increasing the lag between presentations acts only to increase encoding variability. This relationship can be seen simply be noting the positive relationship between lag (t) and performance in that model (as shown in Eq. (2)).

9.3. Memory for associates

In the case of pure repetition, it is presumed that the potential for reminding is very high—that is, a stimulus is a highly effective cue for reminding learners of prior episodes with that stimulus. Conceptualizing the effects of repetition as a consequence of reminding permits generalization to other cases in which items are not repeated but reminding is thought to occur. Here we consider briefly how this model can be applied to cases in which associates or variants of a common stimulus are presented at the two presentations (e.g., Hintzman, Summers, & Block, 1975; Appleton-Knapp et al., 2005; Thios, 1972).

In the simulated data shown in the top panel of Fig. 5, the parameter governing the reminding potential of a second stimulus ($R[p_1|p_2]$) is varied. The dark line shows the curve for a very high value (0.99), representing the effects of repetition. The middle curve (shown in darker gray) represents a reminding value of 0.4, and the lowest curve (shown in the lightest gray) represents a reminding value

Author's personal copy

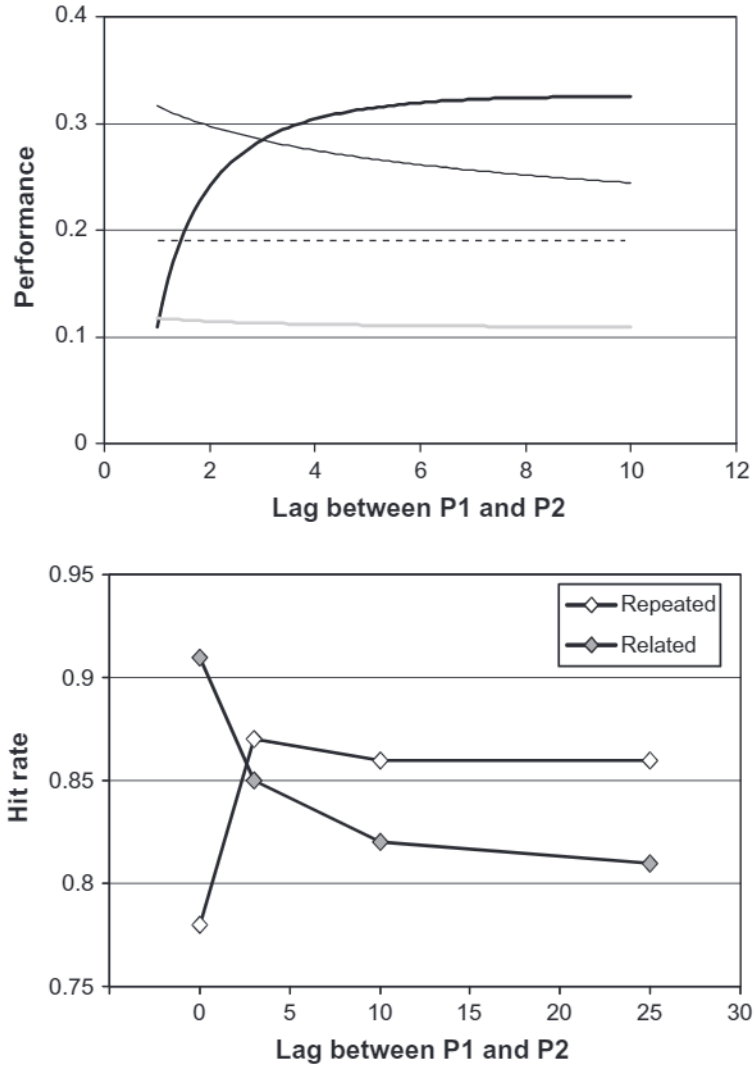


Fig. 5. Top panel: The effects of reminding on repeated items (darkest line), high-associate pairs (next darkest shade), and nonassociates (lightest shade). The superadditivity baseline is indicated by the dashed line. Bottom panel: Data reproduced from Hintzman et al. (1975) comparing recognition of repetitions and associate pairs.

of 0.02. This value can be thought of as similar to the backwards associative strength for the item at p_2 to the item at p_1. Several aspects of the darker gray curve are significant. First, in contrast to the curve for repeated items, it is monotonically decreasing throughout the lag range shown in the figure. In fact, however, both functions are nonmonotonic, but the inflection point for the associate function is earlier than that for the repetition function.

The relative positions of these predicted curves can be compared to the empirical data reported by Hintzman et al. (1975, Table 1), shown in the bottom panel of Fig. 5. This figure shows hit rates on a test of recognition for previously repeated (lighter diamonds) and previously related (i.e., associate) pairs (darker diamonds). These data confirm the central prediction of the model that the inflection point for repetitions is further out in time than the corresponding point for associate pairs.

Another difficulty that theories of encoding variability face is the fact that the lag between unrelated items does not affect memory for those items (Ross & Landauer, 1978). If the benefits of spacing owe to increasing encoding variability, then subjects should be able to recall one of two items presented at a longer lag—and consequently encoded with a greater variety of contextual elements—at a higher rate than if they were presented at a short lag. This prediction was disconfirmed by Ross and Landauer (1978).

Because the relationship between presentations is central to the reminding model, it handles this effect quite straightforwardly, as shown by the lightest line in Fig. 5. When the associative strength and resultant reminding potential between two items is very small (here it is .02), the *RP* term is very

Author's personal copy

close to zero and its effects overwhelm the other factors in Eq. (2). That is, the probability of being reminded of a particular past item is very low, so the value of that reminding is mostly irrelevant.

9.4. RI functions

Earlier we demonstrated that Stimulus Sampling Theory cannot produce nonmonotonic lag functions, and that they can thus not provide an adequate explanation of classic result reported by Glenberg (1976, Fig. 1) and replicated by Cepeda et al. (2008, Fig. 3) that increasing *RI* changed the shape of lag functions by shifting the inflection point further to the left. Encoding variability theory has been thought to be able to account for this result by noting that at short *RI*, contextual elements at test overlap considerably with ones present during the recent study trials; thus, the specificity of the test to two study presentation close in time to one another is greater than for two study presentations presented distantly from one another. At long *RI*, however, when the contextual elements at test are dissimilar from study presentations, the generalizability provided by a more variable encoding context is more beneficial. However, in the version implemented here derived from Stimulus Sampling Theory, encoding variability cannot actually handle these data, as discussed previously.

As presented to this point, the reminding model has no mechanism for incorporating the effects of *RI*, so it must be augmented to order to account for such results. To do so, we draw on the principles of Wickelgren (1972) and assume that repetition leads to retarded forgetting. We further assume that the degree of this retardation is related to the value of reminding (i.e., $1 - RP$):

$$M(p_1|RI = r) = M(p_1)r^{\frac{-k}{\beta(1-RP)+1}}, \tag{6}$$

in which r is the positive retention interval ($r \geqslant 1$), β is a scaling parameter (set for present simulations to be 1), and all other parameters are as defined previously. In this model, the rate at which items are forgotten is lower following a re-presentation or the presentation of a related item, and the degree to which it is lower is proportional to the extent of the mnemonic benefit accorded during that reminding event. Thus, forgetting is slowed to a greater degree when reminding is either temporally or associatively distant.

A simulation with this model is shown in Fig. 6. All important qualitative effects evident in the Glenberg (1976) data are captured here; most importantly, the inflection point at short *RI* is

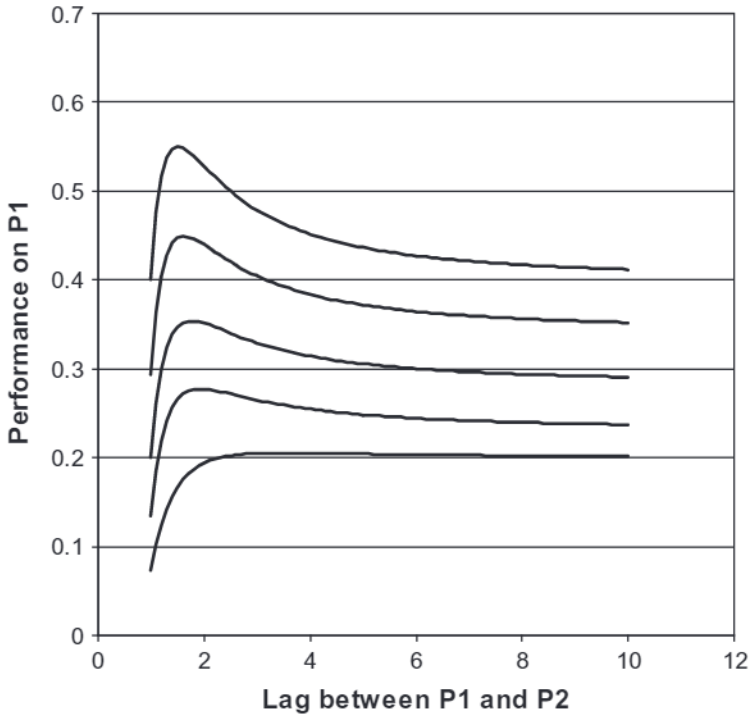


Fig. 6. Simulation of the interaction between lag and retention interval based on Eq. (3).

Author's personal copy

considerably shorter than at longer *RI*. However, it is worth noting that the reminding model can only account for this result with the additional assumption that the effects of forgetting are reduced to a greater degree when the reminding potential is high.

10. What makes distributed practice effective?

We have made two arguments here about the origins of distributed-practice effects. The first is that the concept of encoding variability faces some challenges in dealing with basic and common findings: Superadditive levels of memory performance for repeated items are ubiquitous in the literature and are incompatible with Stimulus Sampling Theory. Though such theories can be modified to handle these effects, they seem to challenge some of the most basic assumptions of encoding variability theory—namely, that the benefits of spacing obtain because greater spacing engenders more variable and thus more generalizable representations. Superadditivity can be added to the list of empirical results, including the failure to obtain spacing effects with unrelated items (Ross & Landauer, 1978), and the finding that failing to remember the first presentation at the time of the second eliminates the effects of spacing (Bellezza, Winkler, & Andrasik, 1975), that suggest that encoding variability theory, as embodied in SST, falls short of capturing important aspects of the empirical body of data on the effects of distributed practice. In addition, encoding variability theory is unable to predict nonmonotonic lag functions, even when the response function is not assumed to be linear, unless the all-or-none element encoding assumption is also abandoned.

Our second argument is that a simpler route to understanding distributed practice is to consider the effects of reminding during study. When a stimulus or task reminds the learner of a previous episode, the retrieval of that episode enhances memory for its contents. Our simple model that implements these assumptions easily handles all of the data that are problematic for encoding variability theory, and with only one additional assumption can account for the effects of retention interval on the shape of the spacing function. Reminding theory also appears to be able to handle many of the other successes of encoding variability theory, such as the effects of a secondary task during the interval between p_1 and p_2 (Bjork & Allen, 1970).

Reminding theory provides a general theoretical stance under which both the effects of repetition and the effects of relationships among different study items can be understood. The heritage of the reminding theory presented here is in the theories of study-phase retrieval (Hintzman, 2010; Thios & D'Agostino, 1976) and consolidation (Wickelgren, 1972), as well as in the idea of retroactive facilitation (Bruce & Weaver, 1973) from the verbal learning tradition. Retroactive facilitation describes the counterintuitive phenomenon that memory for a list of A–B pairs is superior following A–B, A–C than A–B, D–C learning conditions. That phenomenon is counterintuitive because interference theory (Underwood, 1957) clearly predicts the opposite: A–C study should interfere more with memory for the A–B list than should D–C study (but see Wixted, 2004). However, if one considers the fact that members of the A–C list are more likely to induce reminding of corresponding pairs from the A–B list than are members of the D–C list, then retroactive facilitation seems a reasonable consequence of such a study regimen.

11. Implications of reminding theory

Theories of reminding—as opposed to contextual variation—unify research on the effects of repetition with two other major literatures typically not considered as related. We have considered one here: how the effects of varying semantic and associative relationships between studied items are similar to the effects of repetition. Ongoing work in our lab (Tullis, Benjamin, & Ross, in preparation) uses such relationships to explore the form of spacing functions and rate of forgetting for singly presented items that are related to other items in the study list. Such research bears on a theory of reminding, but does not play an obvious role under the encoding variability aegis. Second, because reminding is a concept widely used in research on higher-level cognition, thinking about the basic causes and mnemonic consequences of reminding serves to relate basic memory theory to research on concept learning, problem solving, and understanding (Benjamin & Ross, in press).

Author's personal copy

Reminding theory also suggests that optimal learning is achieved when schedules of acquisition appropriately balance the costs and benefits of forgetting. Too much forgetting leads to unlikely reminding, and too little forgetting leads to impotent reminding. In the formulation provided here, associative similarity operates similarly to forgetting: too little similarity fails to elicit reminding, and too much similarity engenders reminding with little mnemonic benefit.

Because reminding both enhances memory and retards forgetting, there are practical consequences of reminding for the implementation of successful learning regimens. One straightforward implication of the fact that reminding retards forgetting is that expanding-interval retrieval schedules should prove superior for long-term retention than constant-interval schedules. This implication appears to be true (Landauer & Bjork, 1978), though recent research has called its generality into question (Karpicke & Roediger, 2007; Logan & Balota, 2008).

A second implication of the reminding view presented here is that the memory enhancement occurs for the original event, not for the reminding event (though an alternative reminding theory could certainly involve generalization and contrast of those two events, which would enhance memory for both events). Applied to skill learning, this view suggests that it is the quality of the original encoding that is particularly important for successful acquisition. If one initially learns a poor golf swing, for example, then the many later practice opportunities will serve to reinforce those bad habits, not correct them. How feedback plays a role in such corrections remains unexplored: it may be that corrective feedback can decrease reminding on an undesirable original memory, or that it serves to "tune" that memory by enhancing reminding of the original event and of the corrective information.

12. Reminding in extant memory models

Though extant memory models do not generally possess an explicit role for reminding, there are theories for which very similar mechanisms play a critical role. Most prominently, the Temporal Context Model (TCM) of Howard and Kahana (2002) and Howard, Kahana, and Wingfield (2006) utilizes reminding to explain ordering effects in the output of free recall. In their model, recall of an item from a list entails recovery of the context associated with that item, and that context is then used to cue recall of additional items from the list. Because context in their model fluctuates in a very similar manner to the way it does in SST, such a mechanism explains the positive autocorrelation in recall as a function of input serial position, as well as why items close in input position tend to be recalled closely together. A process-based reminding model of spacing based on the principles outlined here generalizes this principle from test-based reminding to reminding during the study phase. Models that contain such mechanisms, sometimes called the "retrieval-dependent update assumption," (Mozer et al., 2009) include those of Mozer et al. (2009), Pavlik and Anderson (2005), and Raaijmakers (2003), though the model of Pavlik and Anderson is the only one that does not augment reminding with a contextual-variability mechanism.

13. Conclusions

Reminding theory will, of course, have to confront its own set of challenges. As it stands, for example, it has no means for dealing with differential effects of spacing on different tests (e.g., Greene, 1989). It also places the locus of the spacing enhancement at p_1, but there is some evidence that memory for both p_1 and p_2 is affected by spacing (Hintzman et al., 1973). These difficulties are not unique to reminding theory, but they do illustrate how reminding needs to be placed within a more global model of encoding and memory access in order to evaluate the theoretical stance more broadly.

A successful theory of distributed practice must have at its core some form of positive interactivity between study events to yield superadditive levels of performance, as reminding theory does. The balance between the probability of reminding and the value of reminding determines, for any task, stimulus, and subject, a "sweet spot" at which spacing is maximally effective. It is our hope that this paper will help spur development of versions of such theories that can be evaluated with the same rigor as the veritable encoding variability theories that have proven useful in advancing understanding in much of human and animal learning and forgetting.

Author's personal copy

Acknowledgments

This research was funded in part by Grant R01 AG026263 from the National Institutes of Health to ASB. We thank Brian Ross, John Wixted, Larry Jacoby, Douglas Hintzman, and the members of the Human Memory and Cognition Laboratory at the University of Illinois at Urbana-Champaign for their helpful insights, comments, and criticism.

Appendix A

In this Appendix, we demonstrate our generalization of Stimulus Sampling Theory (Eq. (1)) to cases in which retention interval is varied using only the parameters of the original model (and one new one representing the retention interval itself), ending up with Eq. (3).

We start by assuming that the encoded elements that are available at the time of test come from one of two sources: they are encoded during p_1 and are also in the available set at test (it is not relevant whether they are available during p_2, according to the all-or-none element encoding assumption), or they are not in the available set during p_1 but are encoded during p_2, and are in the available set for the test. The probability of recall at the time of the test is the sum of the probabilities of these two conditions.

$$R(RI, t, m = 2) = P_{p_1} P_{\text{test}} + P_{outduringp_1} P_{p_2} P_{\text{test}},$$

in which P_{px} indicates the probability of an element being available at presentation x, P_{test} indicates the probability of an element being available at test, and $P_{outduringpx}$ indicates the probability of an element *not* being available at presentation x.

The probability of an element being in the available set during P_x is J, and the probability of an element being out of that set is $(1 - J)$. The probability of an element from the available set also being in the current set following an interval t is specified by SST as:

$$P_{\text{remain}} = J + (1 - J)a^t.$$

and the probability of an element in the unavailable set moving into that set during time period t is specified as:

$$P_{movein} = J(1 - a^t).$$

Putting these terms together and substituting in the various interval parameters (t and RI) yields the general model shown in Eq. (3):

$$R(RI, t, m = 2) = J[J + (1 - J)a^{t+RI}] + (1 - J)[J(1 - a^t)][J + (1 - J)a^{RI}],$$

Appendix B

In this Appendix, we demonstrate that the derivative of Eq. (3) does not yield a value of 0 at any positive retention interval. We start with Eq. (4) and set the left-hand side of the equation to 0:

$$0 = J^2 a^t \ln a(J - 1 - Ja^{RI} + a^{RI}),$$

If, as previously assumed, J and a cannot equal 0 and $-1 < a < 1$, then $J^2 a^t \ln a$ cannot equal 0 and we can eliminate that term from the equation.

$$\begin{aligned} 0 &= J - 1 - Ja^{RI} + a^{RI}, \\ 1 - J &= a^{RI}(1 - J), \\ 1 &= a^{RI}, \\ \ln(1)/\ln a &= RI, \\ 0 &= RI. \end{aligned}$$

The slope of the function is thus only zero when retention interval is zero, indicating that there are no inflection points at positive retention intervals.

Author's personal copy

References

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science, 2*, 396–408.

Appleton-Knapp, S. L., Bjork, R. A., & Wickens, T. D. (2005). Examining the spacing effect in advertising: Encoding variability, retrieval processes, and their interaction. *Journal of Consumer Research, 32*, 266–276.

Bäuml, K. (1997). The list-strength effect: Strength-dependent competition or suppression? *Psychonomic Bulletin and Review, 4*, 260–264.

Begg, I., & Green, C. (1988). Repetition and trace interaction: Superadditivity. *Memory and Cognition, 16*, 232–242.

Belleza, F. S., Winkler, H. B., & Andrasik, F. Jr., (1975). Encoding processes and the spacing effect. *Memory and Cognition, 3*, 451–457.

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

Bjork, R. A. (1988). Retrieval practice and the maintenance of knowledge. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues* (pp. 396–401). Chichester, UK: Wiley.

Bjork, R. A., & Allen, T. W. (1970). The spacing effect: Consolidation or differential encoding? *Journal of Verbal Learning and Verbal Behavior, 9*, 567–572.

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. 2, pp. 35–67). Hillsdale, NJ: Erlbaum.

Bower, G. H. (1972). *Cognition in learning and memory*. Oxford, England: John Wiley & Sons.

Bruce, D., & Weaver, G. E. (1973). Retroactive facilitation in short-term retention of minimally learned paired associates. *Journal of Experimental Psychology, 100*, 9–17.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132*, 352–380.

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning: A temporal ridgeline of optimal retention. *Psychological Science, 19*, 1095–1102.

Challis, B. H. (1993). Spacing effects on cued-memory tests depend upon level of processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 389–396.

Crowder, R. G. (1976). *Principles of learning and memory*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Ebbinghaus, H. (1964). *Memory: A contribution to experimental psychology*. New York: Dover (original work published 1885).

Estes, W. K. (1955a). Statistical theory of spontaneous regression and recovery. *Psychological Review, 62*, 145–154.

Estes, W. K. (1955b). Statistical theory of distributional phenomena in learning. *Psychological Review, 62*, 369–377.

Glanzer, M. (1969). Distance between related words in free recall: Trace of the STS. *Journal of Verbal Learning and Verbal Behavior, 8*, 105–111.

Glenberg, A. M. (1976). Monotonic and nonmonotonic lag effects in paired-associate and recognition memory. *Journal of Verbal Learning and Verbal Behavior, 15*, 1–16.

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.

Haberlandt, K. (1999). *Human memory: Exploration and application*. Boston: Allyn and Bacon.

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). Potomac, MD: Erlbaum.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95*, 528–551.

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

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

Hintzman, D. L., Block, R. A., & Summers, J. J. (1973). Contextual associations and memory for serial position. *Journal of Experimental Psychology, 97*, 220–229.

Hintzman, D. L., & Rogers, M. K. (1973). Spacing effects in picture memory. *Memory and Cognition, 1*, 430–434.

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

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

Howard, M. W., Kahana, M. J., & Wingfield, A. (2006). Aging and contextual binding: Modeling recency and lag-recency effects with the temporal context model. *Psychonomic Bulletin and Review, 13*, 439–445.

Janiszewski, C., Noel, H., & Sawyer, A. G. (2003). A meta-analysis of the spacing effect in verbal learning: Implications for research on advertising repetition and consumer memory. *Journal of Consumer Research, 30*, 138–140.

Karpicke, J. D., & Roediger, H. L. (2007). Expanding retrieval promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 704–719.

Krug, D., Davis, B., & Glover, J. A. (1990). Massed versus distributed repeated reading: A case of forgetting helping recall? *Journal of Educational Psychology, 82*, 366–371.

Landauer, T. K. (1975). Memory without organization: Properties of a model with random storage and undirected retrieval. *Cognitive Psychology, 7*, 495–531.

Landauer, T. K., & Bjork, R. A. (1978). Optimal rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625–632). London: Academic Press.

Lee, T. D., & Genovese, E. D. (1988). Distribution of practice in motor skill acquisition: Learning and performance effects reconsidered. *Research Quarterly for Exercise and Sport, 59*, 277–287.

Logan, J. M., & Balota, D. A. (2008). Expanded vs. equal interval spaced retrieval practice: Exploring different schedules of spacing and retention interval in younger and older adults. *Aging, Neuropsychology, and Cognition, 15*, 257–280.

Author's personal copy

Madigan, S. A. (1969). Intraserial repetition and coding processes in free recall. *Journal of Verbal Learning and Verbal Behavior, 8*, 828–835.
Melton, A. W. (1967). Repetition and retrieval from memory. *Science, 158*, 532.
Melton, A. W. (1970). The Situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behavior, 9*, 596–606.
Mozer, M. C., Pashler, H., Cepeda, N., Lindsey, R., & Vul, E. (2009). Predicting the optimal spacing of study: A multiscale context model of memory. In Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams, & A. Culotta (Eds.). *Advances in neural information processing systems* (Vol. 22, pp. 1321–1329). La Jolla, CA: NIPS Foundation.
Paivio, A. (1974). Spacing of repetitions in the incidental and intentional free recall of pictures and words. *Journal of Verbal Learning and Verbal Behavior, 13*, 497–511.
Pavlik, P. I., & Anderson, J. R. (2005). Practice and forgetting effect on vocabulary memory: An activation-based model of the spacing effect. *Cognitive Science, 29*, 559–586.
Peterson, L. R., Hillner, K., & Saltzman, D. (1962). Time between pairings and short-term retention. *Journal of Experimental Psychology, 64*, 550–551.
Peterson, L. R., Wampler, R., Kirkpatrick, M., & Saltzman, D. (1963). Effect of spacing presentations on retention of a paired associate over short intervals. *Journal of Experimental Psychology, 66*, 206–209.
Raaijmakers, J. G. W. (2003). Spacing and repetition effects in human memory: Application of the SAM model. *Cognitive Science, 27*, 431–452.
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. 1). New York: Academic Press.
Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88*, 93–134.
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.
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.
Russo, R., Parkin, A. J., Taylor, S. R., & Wilks, J. (1998). Revising current two-process accounts of spacing effects in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 161–172.
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). Oxford: Oxford University Press.
Sternberg, R. J. (2003). *Cognitive psychology*. Belmont: Thomson Wadsworth.
Thios, S. J. (1972). Memory for words in repeated sentences. *Journal of Verbal Learning and Verbal Behavior, 11*, 789–793.
Thios, S. J., & D'Agostino, P. R. (1976). Effects of repetition as a function of study-phase retrieval. *Journal of Verbal Learning and Verbal Behavior, 15*, 529–536.
Tullis, J. G., Benjamin, A. S., & Ross, B. H. (in preparation). *The role of reminding in memory tasks*.
Underwood, B. J. (1957). Interference and forgetting. *Psychological Review, 64*, 49–60.
Watkins, M. J., & Kerkar, S. P. (1985). Recall of twice-presented item without recall of either presentation: Generic memory for events. *Journal of Memory and Language, 24*, 666–678.
Waugh, N. C. (1963). Immediate memory as a function of repetition. *Journal of Verbal Learning and Verbal Behavior., 2*, 107–112.
Wickelgren, W. A. (1972). Trace resistance and the decay of long-term memory. *Journal of Mathematical Psychology, 9*, 418–455.
Wixted, J. T. (2004). On common ground: Jost's (1897) law of forgetting and Ribot's (1881) law of retrograde amnesia. *Psychological Review, 111*, 864–879.
Wixted, J. T., & Ebbesen, E. B. (1991). On the form of forgetting. *Psychological Science, 2*, 409–415.