ELSEVIER

Cognitive Science 27 (2003) 431–452

COGNITIVE SCIENCE

http://www.elsevier.com/locate/cogsci

Spacing and repetition effects in human memory: application of the SAM model

Jeroen G.W. Raaijmakers*

Department of Psychology, University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands

Received 10 April 2002; received in revised form 11 November 2002; accepted 28 December 2002

Abstract

Spacing between study trials of an item increases the probability that item will be recalled. This article presents a new model for spacing based on the SAM theory of memory developed by Raaijmakers and Shiffrin (1980, 1981). The model is a generalization of the SAM model as applied to interference paradigms (Mensink & Raaijmakers, 1988, 1989) and may be viewed as a mathematical version of the Component-Levels theory proposed by Glenberg (1979). It is assumed that on a second presentation of an item, information is added to an existing trace if the episodic memory image corresponding to that item is retrieved. If it is not retrieved, a new image is stored. It is shown that the model predicts many standard findings including the lack of a spacing effect for the recall of at least one of two items each presented once (Ross & Landauer, 1978).

Keywords: Learning paradigms; Episodic memory image; Standard findings

1. Introduction

That performance in a memory task increases with repetition of the material is a basic and somewhat trivial finding in memory research. What is not trivial however is the question why. What exactly happens when an item is presented for a second time? Do we store two memory traces, one for each occurrence, or do we somehow strengthen the trace that was formed on the first trial? Over the years this question has been answered in many different ways. Some theories hold that each presentation is stored separately (e.g., Landauer, 1975) whereas others assume

* Tel.: +31-205-256-839; fax: +31-205-391-656.
E-mail address: raaijmakers@psy.uva.nl (J.G.W. Raaijmakers).

doi:10.1016/S0364-0213(03)00007-7

only a single memory trace where additional presentations either increase the probability of storing the trace in long-term memory (e.g., models of the all-or-none type, Bower, 1961), or increase the strength of the trace.

A related issue concerns the effects of spacing of individual repetitions. A well-known phenomenon that has been observed in many learning paradigms is the *distributed practice* or *spacing effect*. As a general rule, provided the retention interval is not too short, the probability that a person will recall an item that has been presented twice, increases as the interval or lag between the two presentations is increased (and similarly for more than two presentations). This finding is intriguing since the spacing effect occurs despite the fact that the retention interval for the first presentation is larger in the spaced presentation situation and hence one might have expected more forgetting of the corresponding memory trace rather than the increase that is in fact observed.

An additional complication for any model that claims to provide a complete account of spacing effects, is that the effect interacts with the length of the retention interval: at short retention intervals the effect reverses and performance decreases with an increase in the spacing interval. Moreover, at intermediate retention intervals, the effect is non-monotonic. This latter finding has been especially troublesome for many accounts for spacing effects.

The spacing effect is a robust finding that is not only obtained with paired associates or other typical laboratory material but also in real-life training and learning situations (Baddeley & Longman, 1978; Bahrick & Phelps, 1987; Smith & Rothkopf, 1984). This probably means that basic principles of learning and retention are involved. As we will show in this article, certain findings in this literature also have implications for our initial question concerning separate versus cumulative memory traces.

In this article we will show that the SAM theory for recall as developed by Raaijmakers and Shiffrin (1980, 1981) and incorporating the contextual fluctuation extension developed by Mensink and Raaijmakers (1988, 1989), gives a good account for spacing effects. The importance of this demonstration is that it shows that the SAM theory provides a relatively complete explanation for all of the major phenomena of episodic recall. When SAM was originally introduced, the goal was to show that the results from a variety of memory paradigms could be fitted within a single theoretical framework. Such a goal was quite ambitious at the time (and probably still is) but, 20 years after its introduction, it seems a fair conclusion that SAM has been relatively successful. Although there have been a number of problems (especially with respect to the recognition data), the model still stands as one of the major frameworks for episodic memory.

Recently, a new model has been developed by Shiffrin and Steyvers (1997). This model, named REM, was developed initially to deal with the problematic results in recognition memory but has since been extended to recall paradigms as well as to some implicit and semantic memory paradigms (see Raaijmakers & Shiffrin, 2002; Schooler, Shiffrin, & Raaijmakers, 2001). Thus, REM is an even more ambitious attempt at developing a global memory theory since it is not restricted to episodic memory paradigms as was the case with SAM. Interestingly however, the REM model for recall is formally similar to the SAM recall model and hence the previously obtained results for recall can be directly generalized to REM. In this article however we will restrict ourselves to SAM (primarily because the simulations that will be reported were in fact based on SAM rather than REM).

We will first give a brief summary of the general SAM theory for recall. Next, we will give a brief summary of one important theoretical account of spacing effects, namely Glenberg's Component-Levels theory (Glenberg, 1979). Following that, we will describe in detail the SAM model for spacing and repetition effects and how it handles the basic phenomena in this area.

2. A brief summary of SAM

SAM (Search of Associative Memory) was originally developed as a model for free recall (see Raaijmakers, 1979). However, it was quickly recognized that the general framework of SAM could be applied not just to free recall but also to other memory paradigms such as paired-associate recall and recognition (see Gillund & Shiffrin, 1984). The SAM theory is a probabilistic cue-dependent search theory that describes retrieval processes in long-term memory. The basic framework of SAM assumes that during storage, information is represented in memory traces or "memory images," that contain item, associative and contextual information. The amount and type of information stored is determined by coding processes in STS (elaborative rehearsal). In SAM, retrieval is assumed to be based on retrieval cues such as experimenter-provided words, category names, self-generated words from a to-be-remembered list and contextual cues.

Whether an image is retrieved or not, depends on the associative strengths of the retrieval cues to that image (these strengths are usually assumed to be proportional to the length of time an item has been studied). SAM incorporates a rule to compute the overall strength of a set of probe cues to a particular image: let $S(Q_j, I_i)$ be the strength of association between cue Q_j and image I_i. Then the combined strength or activation of image I_i, $V(i)$, for a probe set consisting of $Q_1, Q_2, \ldots, Q_m$ is given by

$$V(i) = \prod_{j=1}^{m} S(Q_j, I_i) \tag{1}$$

An important property of this multiplicative rule is that it allows the search process to be focused on those traces that are associated to all cues rather than those that are associated to just one of the cues. That is, those traces that are not associated to one of the cues ($S(Q_j, I_i) = 0$ for some j) will have a combined activation equal to 0.

In paired-associate recall tasks (the paradigm that will be used throughout this paper), the images stored in memory correspond to the word pairs on a given list (items of the form *A*–*B*) and the cues that are used during retrieval are the stimulus member of the pair (the *A* member of the *A*–*B* pair) and the contextual cues present at the time of testing. The search process consists of a series of retrieval attempts using the same set of cues. Each attempt involves selecting or sampling one image based on the activation strengths V_i. The probability of sampling image I_i equals the relative strength of that image compared to all images in LTS:

$$P_S(I_i) = \frac{V(i)}{\sum V(k)} \tag{2}$$

Sampling an image allows recovery of some of the information from it. For simple recall tasks, the probability of successfully recovering the name of the encoded word after sampling the image I_i is assumed to be an exponential function of the sum of the strengths of the probe set to the sampled image:

$$P_R(I_i) = 1 - \exp\left[-\sum_{j=1}^{m} S(Q_j, I_i)\right] \quad (3)$$

If the retrieval is not successful, a new retrieval attempt is made and this continues until a criterion of $L_{\max}$ failures have occurred. Thus, the probability of recall is given by the probability that the item was sampled at least once, times the probability that recovery was successful:

$$P_{\text{recall}}(I_i) = [1 - (1 - P_S(I_i))^{L_{\max}}]P_R(I_i) \quad (4)$$

Note that it is assumed here that whenever an incorrect trace (say corresponding to a pair *C–D*) is sampled and recovered (a relatively rare event due to the low associative strength between the cue *A* and the trace *C–D*), the recovered information will allow the rejection of that item (i.e., it is recognized that the recovered information does not match the cue *A*). Although this assumption is of course a bit too strong since it does not allow for intrusions to occur, it has little or no effect on the predicted recall probabilities since the probability of successful recovery of an incorrect trace is very low and such a trial would have nearly always resulted in an error anyway. However, this assumption has the advantage of simplifying the model since it enables the derivation of closed expressions for the predicted probability of recall as in Eq. (4).

Raaijmakers and Shiffrin (1980, 1981) showed that a model based on these assumptions could explain many basic results in free and cued recall. These included serial position effects, the effects of list length and presentation time, the effects of cuing with category names in categorized free recall, cumulative recall curves and inter-response times in free recall, hypermnesia effects in repeated recall, and part-list cuing effects (see also Raaijmakers & Phaf, 1999) for additional evidence for SAM's explanation of part-list cuing). Although many of these predictions were not very surprising given the basic structure of the model (and its similarity to previous models proposed by Atkinson & Shiffrin, 1968; Shiffrin, 1970), some of the predictions were novel and indeed counterintuitive (see e.g., Raaijmakers & Phaf, 1999).

Despite its successes, the original SAM model was not equipped to handle basic forgetting effects (other than through increased competition). In order to be able to handle time-dependent changes in recall, Mensink and Raaijmakers (1988, 1989) proposed an extension of the SAM model, the contextual fluctuation model. The basic idea, adapted from Stimulus Sampling Theory (Estes, 1955), is that a random fluctuation of elements occurs between two sets, a set of available context elements and a set of (temporarily) unavailable context elements. The contextual strengths at test are a function of the relationship between the sets of available elements at study and test. Mensink and Raaijmakers (1989) showed how some simple assumptions concerning the fluctuation process yield equations for computing the probability that any given element is active both at the time of storage and at the time of retrieval. A more elaborate analysis of contextual fluctuation processes and its application to free recall was recently proposed by Howard and Kahana (1999), see also Kahana (1996).

Using this contextual fluctuation version of SAM, it was shown that a model could be developed that was capable of providing a unified account of the interference and forgetting phenomena of the classical interference literature. These included the basic results concerning retroactive inhibition, proactive inhibition, spontaneous recovery, independence of List 1 and List 2 recall, Osgood's transfer and retroaction surface, as well as simple forgetting functions (i.e., single-list paradigms). Mensink and Raaijmakers (1988) showed that these phenomena could be explained by SAM without the need for an "unlearning" assumption, an assumption that was shown to be the major reason for the difficulties encountered by traditional interference theories in trying to provide a unified theory for forgetting.

In the next sections, we will show that this contextual fluctuation model also provides a relatively simple explanation for spacing effects. However, before addressing the way in which SAM may provide an explanation for spacing effects, we will first describe a non-quantitative theory for spacing effects proposed by Glenberg (1979) that shares many features with the SAM approach and has been a source of inspiration in our modeling attempts.

3. Glenberg's Component-Levels theory

In the 1970s many theories were proposed to account for spacing and repetition effects (see Crowder, 1976; Hintzman, 1974, 1976 for detailed reviews), usually based on either some notion of consolidation (e.g., Landauer, 1967, 1969) or encoding variability (e.g., Madigan, 1969; Melton, 1970). The most complete explanation is probably the *Component-Levels theory* proposed by Glenberg (1979). This theory assumes that a stimulus is represented by a multi-component episodic trace. Which components (or features) are included in a trace, depends on the actual stimulus that is presented, the nature of the processing task, the subject's strategies and the context in which the stimulus is presented (Glenberg, 1979, p. 96). Glenberg (1979) distinguishes three types of components: *contextual* (representing the context at presentation), *structural* (relations and associations between items), and *descriptive* (specific item features). These components differ as to the probability that they are included in traces representing different items and the probability that they vary between successive presentations of the same item. Contextual components are automatically included in all traces presented in the same context. Context is assumed, however, to drift over time, leading to variability between successive presentations of the same item. The structural components that are encoded in the trace, are less general and depend on the other items being processed simultaneously. Storage of these components is not automatic but depends on control processes used by the subject and is influenced by such factors as the nature of the task and the task instructions. Still more specific are the descriptive components. These are copied from a semantic memory representation into the episodic memory trace. They do however depend on the nature of the processing in which the subject is engaged and the context (e.g., depth of encoding).

Generally speaking, spacing of presentations will lead to more contextual, structural, and descriptive components being stored in the memory trace. However, performance (and hence the spacing effect) is not only determined by the components that are encoded in the trace, but also by the cues used at the time of retrieval. These retrieval cues activate corresponding components in the memory trace. The degree of activation of a component is inversely related to the number of traces in which that component is included (its generality). The degree of activation of a trace (apparently assumed to be directly related to memory performance) is a monotonic function of the summed activation of its individual components. These assumptions lead to three corollaries: (1) trace activation and retrieval are functions of the number of components shared by the cue and the trace, (2) trace activation decreases as the generality of the components in the trace or the cue increases, and (3) in general, trace activation increases with the number of components included in the trace (Glenberg, 1979, p. 98). Glenberg (1979) shows how this conceptual framework can be used to explain a variety of spacing and lag effects in various recall and recognition paradigms.

It is not very difficult to see that Glenberg's Component-Levels theory is in many aspects quite similar (on a non-quantitative level) to SAM and especially the extension proposed by Mensink and Raaijmakers (1988, 1989) that includes the assumption of contextual fluctuation. The SAM model assumes that sampling of a trace is a function of the strength of the association of the cues to that trace, relative to their associative strength to other traces. Once a trace has been sampled, its recovery depends on the absolute strength of the trace (which would be similar to the overlap in features between cue and trace). The nature of the sampling rule (the product rule) also predicts that the most specific cue will mostly determine which trace will be accessed. The assumption that features vary in their generality and that this affects the retrieval probability, is also consistent with the SAM theory. Finally, both theories assume that contextual components are subject to a gradual fluctuation process that leads to greater storage with longer spacing intervals.

In this article, we will show that a SAM model may indeed be developed that provides a good quantitative explanation of spacing phenomena. Given the obvious qualitative similarity between the SAM theory and Glenberg's Component-Levels theory, it follows that this model may also be viewed as a quantitative version of Glenberg's theory. We will show that this model not only explains the major spacing and repetition phenomena but also an intriguing finding presented by Ross and Landauer (1978) as inconsistent with any sort of encoding variability model (a conclusion apparently shared by Glenberg & Smith, 1981). We will first describe the spacing model in detail, followed by a comparison to data from several classic experiments. We will then analyze the Ross–Landauer result. We will conclude with a brief discussion of the merits of this particular model for spacing and a few aspects that might be problematic.

4. The SAM model for spacing and repetitions effects

In a typical spacing experiment, the subject might be given two presentations of a paired-associate item *A*–*B* (P_1 and P_2) followed by a test trial in which the stimulus member of the pair (*A*) is given and the subject has to generate or recall the paired member (*B*). Such an experiment might be schematically represented as

$$P_1 \xrightarrow{t_1} \overset{t_p}{P_2} \xrightarrow{t_2} T$$

where t_1 represents the interpresentation or *spacing interval*, t_p equals the presentation time of an item, and t_2 is the *retention interval*. The basic spacing effect refers to the finding that when the retention interval is reasonably long the probability of recalling *B increases* with the length of the spacing interval (t_2 is kept constant). Such an increase is remarkable since a simple model in which each of the two presentations would be stored separately would predict a decrease in the probability of recall since the retention interval for the first presentation ($=t_1 + t_p + t_2$) obviously increases as t_1 increases and hence there should be more forgetting with spaced presentations.

The model that we will be using in this application is the same as the one used by Mensink and Raaijmakers (1988, 1989) except that additional assumptions will have to be made concerning what happens when a particular item is repeated after some interval. In particular, when a new item is presented, it enters a short-term buffer (with probability w) and a new trace is formed in LTS. At a second presentation one of three things may happen:

1. the item is still in the short-term buffer, in which case the response will be correct and no new information will be stored in LTS,
2. the item is in LTS and can be retrieved, in which case the response will also be correct and new information is added to the original trace, or
3. the item is in LTS and cannot be retrieved, in which case the response will not be correct (unless guessing is possible) and a new trace will be formed in LTS.

In cases where there are more than two study trials or presentations, the same rules are used. That is, information is added to the existing trace if the item is successfully retrieved, otherwise a new trace is formed. An important question is what happens to a trace that could not be retrieved or recovered at a certain trial. Would it still be possible to retrieve such a trace at a later test? This is a tricky issue. On the one hand it would seem to be perfectly consistent with SAM to allow for such a possibility. On the other hand, there are a number of factors that make such an event unlikely. First, the SAM model includes the assumption that when a trace is sampled but not recovered, recovery will also fail the next time that item is sampled using the same set of cues. Gronlund and Shiffrin (1986, p. 558) modified this assumption in such a way that the recovery probability was only based on those cues that had not been involved in earlier, unsuccessful retrieval attempts. If such an assumption would be generalized to the present case, it would imply that for a trace that was not retrieved on a given trial, the recovery probability would be based on only those contextual elements that were not present at the time of the initial testing. Second, although the present formulation of the model does not incorporate variability in the associative strengths, such variability must obviously be present. Ignoring this aspect works fine in most cases, except of course when one looks at conditional probabilities such as the probability of successful retrieval given no success at an earlier trial. In such a case, variability has to be taken into account and would lead to a decrease in the conditional probability of retrieval. Third, the probability of retrieving the old trace would be even lower than at the first test since there would be an increase in the retention interval as well as an additional trace that would be interfering with its retrieval (i.e., the new trace formed at the second presentation). All in all then, the probability of retrieving such an old trace would have to be quite low. Empirically, this would be consistent with the results that have been observed with repeated testing in RTT paradigms where a single study trial is followed by two test trials without feedback (see Estes, 1960). In such experiments the usual outcome is that the probability of a correct response on the second test given no success on the first is close to zero (corrected for guessing).

In the present model we therefore made the simplifying assumption that the probability of retrieval was equal to zero for any trace that was not retrieved at a certain trial. Obviously, this assumption cannot be completely correct but for the reasons described above it will probably not make much of a difference. We verified this by calculating the probability of recall for a model in which variability was assumed where we compared the predicted probabilities with or without the inclusion of the old, non-retrieved trace. These probabilities differed by no more than 0.5%. The present assumption has the advantage that it leads to relatively simple analytic solutions for the probabilities of recall.

As in Mensink and Raaijmakers (1988), it is assumed that there is a continuous contextual fluctuation process in which over time the current context becomes more and more dissimilar to the original study context. Since the retrieval strength of the context cue to a memory image is based on the overlap of the current context with the context stored in the trace, this implies that the probability of successful retrieval from LTS will gradually decrease as the retention interval increases.

Assuming that the search continues until $L_{\max}$ failures have occurred, the probability of recall for a retention interval of t seconds is given by (see Eq. (4)):

$$P_{\text{recall}}(t) = [1 - (1 - P_S(t))^{L_{\max}}]P_R(t) \tag{5}$$

where $P_S(t)$ denotes the probability of sampling the correct item for a retention interval of t seconds and $P_R(t)$ is the corresponding probability of successful recovery:

$$P_S(t) = \frac{c(t)I(t)}{c(t)I(t) + Z(t)} \tag{6}$$

and

$$P_R(t) = 1 - \exp[-\theta(c(t) + I(t))] \tag{7}$$

where $c(t)$ is the contextual strength after a retention interval of t seconds and $I(t)$ the corresponding interitem strength between the A cue and the A–B image. The parameter $Z(t)$ in the denominator of Eq. (6) represents the interfering effect of all other associations (other, unrelated, pairs on the list as well as extra-experimental associations: see Mensink & Raaijmakers, 1988, p. 438). In this paper it will be assumed that both $I(t)$ and $Z(t)$ are constant, i.e., they do not depend on the retention interval t. For $Z(t)$ this is of course a simplifying assumption since it does not take into account the effects of new storage for other items during the retention interval. However, since these effects are correlated with the effects due to context fluctuation it is very difficult to separate the two. Moreover, in most experiments the items are embedded in a long list and the retention intervals are relatively short so that the additional effects of new storage for other items will not make much difference.

The parameter θ is introduced to handle the situation where the second presentation is not itself a test trial but only a study trial. In such a case, we assume that an implicit retrieval attempt is made using the presented word pair A–B (plus context) as cues (*study-phase retrieval*). Thus, in the present model it is important whether or not the item is retrieved or recognized at the second presentation, even if no explicit testing occurs. However, it seems reasonable to assume that the probability of successful retrieval for such a case should be higher than when only the A item is present as a cue. Hence, we might assume separate θ values for these two cases (explicit testing and implicit study-phase retrieval). Since these two θ values cannot be separately estimated (due to the confounding with the scaling of the standard strength values), we will fix the θ value for recognition or study-phase retrieval at 1.0 and let the other θ value (for the explicit testing case) free to vary between 0 and 1. These θ parameters are only used in the case where the second presentation is not itself a standard test trial. If all presentations are also test trials, the same θ value applies to all trials (in that case θ is fixed at 0.5).

The most important aspect of the model is how $c(t)$ depends on the length of the interval between two presentations or between the last presentation and the test. It is assumed that $c(t)$ is proportional to the number of contextual elements present at the time of retrieval that were also present at the initial presentation (denoted by $A(t)$). In the simple version that we will use here, it is assumed that all elements that are present on a given study trial will be stored in the memory trace of that item. As shown by Mensink and Raaijmakers (1988, 1989), the number of stored contextual elements that are active after a retention interval of t seconds that were also active at the study trial, $A(t)$, is proportional to

$$A(t) = A(0)\mathrm{e}^{-\alpha t} + Ks(1 - \mathrm{e}^{-\alpha t}) \tag{8}$$

where $A(0)$ is the number of elements that are active at $t = 0$ and K is the total number of contextual elements (active and non-active) stored in the trace. s and α are two parameters that are related to the rate at which contextual elements fluctuate between the active and inactive state. Eq. (8) shows that the number of elements active after a retention interval of t seconds is determined by the number of active elements at the start of the retention interval ($A(0)$), the total number of stored contextual elements (K), as well as the rate of fluctuation between the active and inactive state (s and α). After the first study trial for the item, both $A(0)$ and K are equal to the number of elements active at $t = 0$. Since $c(t)$ is proportional to $A(t)$ and since $A(t)$ itself has a scaling parameter (the $A(t)$ values are all proportional to the absolute number of elements that are active at any given moment) we may set $A(0)$ to a fixed value without affecting in any way the predictions of the model. In the present applications of the model we have therefore set $A(0)$ equal to 1.0.

With more than one study trial followed by a test the situation becomes slightly more complicated. As mentioned above, when the item is recognized (retrieved) at P_2, additional context elements will be stored in the existing trace. It is assumed (as in Mensink & Raaijmakers, 1988) that a given element can be stored only once in a particular trace. It follows (see Mensink & Raaijmakers, 1989) that the number of stored elements active at T after two presentations with a spacing interval of t_1 and a retention interval of t_2 is equal to:

$$A_2(t_1, t_2) = A(t_1, 0)\mathrm{e}^{-\alpha t_2} + K_2(t_1)s(1 - \mathrm{e}^{-\alpha t_2}) \tag{9}$$

where $K_2(t_1)$ equals the total number of contextual elements stored after two study trials with a spacing of t_1 seconds. Note that the notation becomes a bit cumbersome so a brief clarification might be helpful. Eq. (9) is equivalent to Eq. (8) but taking into account that there now have been two presentations (hence the subscript 2) with a spacing interval equal to t_1 seconds. K_2 depends on t_1 since if the spacing interval is short, the two contexts will be more similar and hence there will be fewer contextual elements stored in the combined trace. A_2, the number of elements active t_2 seconds after the second presentation obviously depends on the retention interval t_2 but also on t_1 since it is a function of K_2.

From the assumptions described above, it follows that $A_2(t_1, 0) = A(0)$ and $K_2(t_1)$ is equal to

$$K_2(t_1) = 2A(0) - A(t_1) \tag{10}$$

i.e., $K_2(t_1)$ equals the sum of the number of elements active at P_1 and P_2 minus the number of elements active at both P_1 and P_2 (i.e., the overlap between the context at P_1 and P_2).

Since the contextual strengths are proportional to the overlap between the context at the time of testing and the context elements stored in the trace, we finally get for the context strengths:

for retrieval at P_2:

$$c(t_1) = a\,A(t_1) \tag{11}$$

for retrieval at T given successful retrieval at P_2:

$$c(t_1, t_2) = a\,A_2(t_1, t_2) \tag{12}$$

for retrieval at T given no successful retrieval at P_2:

$$c(t_2) = a\,A(t_2) \tag{13}$$

where a is a proportionality constant (see Mensink & Raaijmakers, 1988, 1989).

As in previous applications of SAM, it is assumed that the interitem strengths are incremented following successful retrieval. In the present model the interitem strength after a single study trial is equal to b, and the corresponding strength after study trial P_2 given successful study-phase retrieval is equal to $b + b_2$ where b_2 represents the effects of incrementing as well as the additional storage due to study on P_2.

For more complicated paradigms involving more than two study trials or presentations, the same logic is used. That is, information is added to the existing trace when the item is successfully retrieved, otherwise a new trace is formed. Finally, the probability that an item is still in the short-term buffer after a delay of t seconds, is simply assumed to be equal to:

$$P_{\mathrm{STS}}(t) = \mathrm{e}^{-\lambda t} \tag{14}$$

In sum, the model is based on the Mensink and Raaijmakers contextual fluctuation model but includes the additional assumption that on a second presentation new information may be added to the previously stored trace but only if that trace is retrieved from episodic or long-term memory. Otherwise a new trace will be formed. In the next sections we will show that this model accounts nicely for the major data on spacing and repetition effects.

5. Fits to standard data on spacing

In this section we will compare the present model to the data from a number of older experiments that have provided reliable parametric data (i.e., that varied one or more design parameters in a parametric way) and that have played an important role in previous theoretical discussions of spacing effects. For ease of understanding we list in Table 1 the parameters of the present model as well the values that were used to generate the predicted data for these experiments.

Table 1
Parameters and their values in the fits to various data sets

Parameters and their meaning		Rumelhart data	Young data	Glenberg data
α	Fluctuation parameter (=sum of rates with which active and inactive elements become inactive and active, respectively)	0.087	0.082	0.013
s	Fluctuation parameter (=ratio of the number of active elements to the total number of context elements)	0.288	0.150	0.260
a	Scaling constant for context association	5.0[a]	5.0[a]	5.0[a]
w	Probability that a new item enters the STS buffer	0.766	1.0[a]	1.0[a]
b	Amount of interitem information stored on a single study trial	0.688	0.246	0.732
Z	Constant representing the interfering effect of other memory traces in sampling	3.0	2.0	10.0
θ_2	Scaling parameter in recovery equation for a test trial	0.5[a]	0.300	0.215
λ	Rate of decay from STS	0.310	0.746	0.800
L_{max}	Maximum number of retrieval attempts	3[a]	3[a]	3[a]

[a] This parameter was not varied but kept fixed in the fitting of the model.

The first set of data that we will fit the model to comes from an experiment performed by Rumelhart (1967, Experiment I). In this experiment, a continuous paired-associate paradigm was used in which the subjects were presented a long continuous list composed of 66 different items (plus filler items) that were each repeated six times with interpresentation lags varying between 1 and 10. Eight different lag sequences were used, each repeated six times throughout the list. The items in this experiment were consonant–vowel–consonant (CVC) trigrams (e.g., KIG, VUP) paired with one of three digits (3, 5 or 7). In this experiment an anticipation procedure was used in which the stimulus item was first presented for test and then, after the subject had responded, the whole pair was shown for an additional 2 s. After the study phase there was a three second inter-trial interval until the next test item was presented. Since there were 50 subjects, each data point was based on 300 observations. Hence, this provides a dataset that is very well suited for modeling purposes.

Table 2 and Fig. 1 give the predicted and observed data for each of the eight conditions of this experiment. The predictions were generated using the model described above including a correction for guessing correctly (equal to 1/3). The parameters were estimated using a general purpose minimization program (MINUIT, see James & Roos, 1975), minimizing the following χ^2-value (a simple algebraic rewrite of the standard Pearson χ^2):

$$\chi^2 = \sum \frac{N(p_o - p_e)^2}{(p_e - p_e^2)} \tag{15}$$

where N is the number of observations and p_o and p_e are the observed and predicted probabilities of correct recall and the summation is over all data points.

Since the present SAM model has a relatively large number of parameters (not all of which may be estimable in a particular application), we have set some of the parameters to fixed values

Table 2
Observed and predicted proportions correct for each of the conditions of the Rumelhart (1967) experiment

Spacing condition	Trial 1		Trial 2		Trial 3		Trial 4		Trial 5		Trial 6	
	OBS	PRE	OBS	PRE	OBS	PRE	OBS	PRE	OBS	PRE	OBS	PRE
10-10-10-10-10	0.320	0.333	0.660	0.677	0.787	0.783	0.847	0.841	0.873	0.878	0.893	0.905
10-10-1-1-10	0.343	0.333	0.703	0.677	0.797	0.783	0.943	0.916	0.967	0.952	0.887	0.895
6-6-6-6-10	0.373	0.333	0.733	0.729	0.810	0.824	0.867	0.872	0.920	0.904	0.907	0.909
6-6-1-1-10	0.370	0.333	0.660	0.729	0.833	0.824	0.917	0.923	0.943	0.953	0.890	0.900
3-3-3-3-10	0.290	0.333	0.793	0.779	0.863	0.868	0.893	0.899	0.917	0.919	0.870	0.897
1-1-10-10-10	0.333	0.333	0.850	0.821	0.907	0.922	0.777	0.788	0.850	0.836	0.880	0.875
1-1-6-6-10	0.310	0.333	0.793	0.821	0.920	0.922	0.840	0.840	0.883	0.868	0.877	0.879
1-1-1-1-10	0.293	0.333	0.850	0.821	0.917	0.922	0.967	0.942	0.973	0.947	0.890	0.853

for all applications. Thus, L_{max} was fixed at 3.0, a was fixed at 5.0 (a value that allows reasonable variation in the context strengths), and b_2 was always equal to b (thus the interitem strength after k presentations was simply equal to $k \cdot b$). In addition, the parameters were restricted to specific ranges in order to avoid unreasonable variation in the parameter values obtained for different experiments (this does have an effect on the goodness-of-fit values but it is only minor). For the Rumelhart experiment, the following values were obtained for the remaining six parameters: $s = 0.288$, $\lambda = 0.310$, $\alpha = 0.087$, $b = b_2 = 0.688$, $Z = 3.00$, and $w = 0.766$ (as mentioned earlier, all θ values were kept at 0.5 since each presentation is also a test trial). The fit of the model is obviously quite good as indicated by the χ^2-value of 38.01 ($df = 34$; if the data from the first guessing trial are also taken into account as in Rumelhart (1967): $\chi^2 = 47.76$, $df = 42$). This is not surprising since the present model is quite similar to the Markov model that was used by Rumelhart (1967) and for which he obtained a similar goodness-of-fit.

What is important in these data is the clear evidence for an interaction effect between the length of the spacing interval and the retention interval. For example, when one compares performance on Trial 4 for the 1-1-10-10-10 condition and the 10-10-10-10-10 condition, it is clear that for a fairly long retention interval of 10, performance on those items that began with a spacing of 1-1 (massed presentation) was clearly worse than the performance on the items that started with spaced presentation (10-10): 0.777 versus 0.847 (predicted values: 0.788 vs. 0.841). However, with a short retention interval massed presentation is superior to spaced presentation. For example, comparing again performance on Trial 4 for the conditions 1-1-1-1-10 and 10-10-1-1-10, the performance for the massed condition is 0.967 while the performance in the spaced condition is 0.943 (predicted values: 0.942 vs. 0.916). This confirms the general finding that spaced presentation is superior to massed presentation except when the retention interval is very short.

In the present model there are two main reasons for the advantage of spaced presentations with a moderately long retention interval such as in this experiment. The first is that a short spacing interval keeps the item in STS and hence prevents the strengthening of the LTS trace. The second factor is that a larger spacing interval leads to more new contextual elements being stored in the trace (provided of course that the original trace is in fact retrieved). Both factors

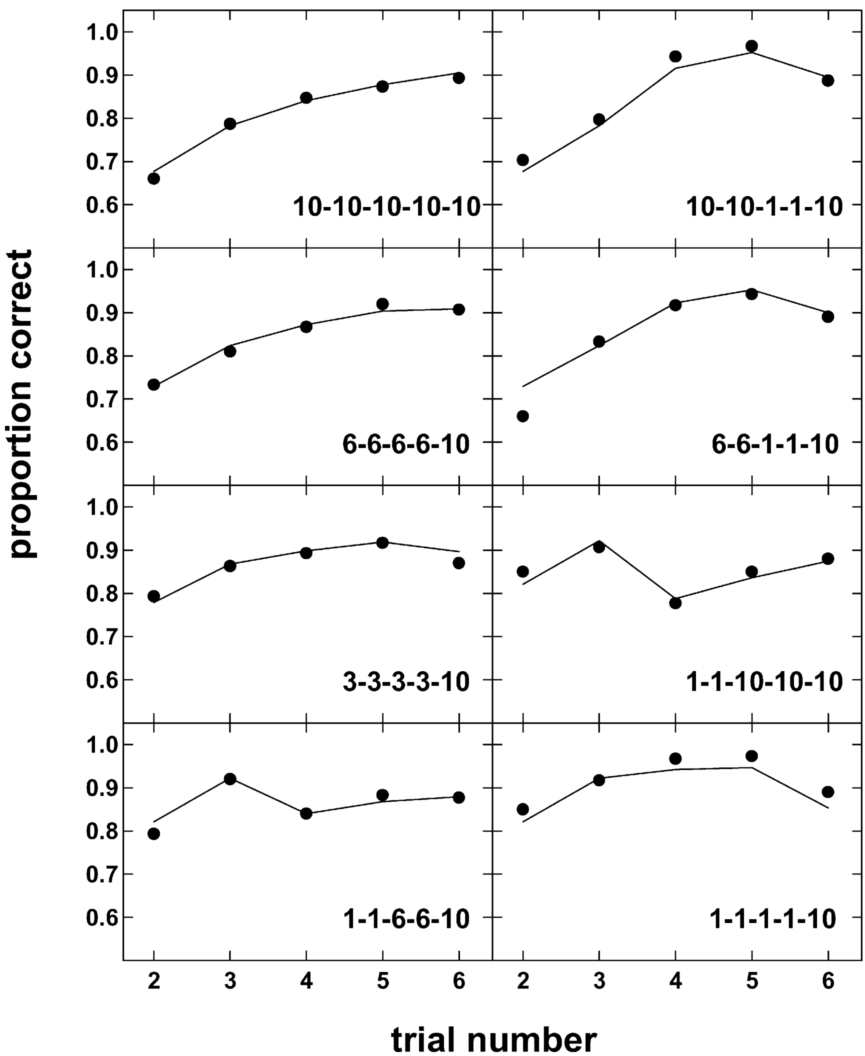

Fig. 1. Observed (dots) and predicted (lines) probabilities of recall for Experiment I of Rumelhart (1967). Each graph gives the spacing (number of intervening items) between successive presentations.

are contributing to the predicted spacing effect and there is no simple way to determine their relative importance.

However, it is not the case that (for longer retention intervals) the probability of retrieval increases monotonically with increasing lag between successive presentations. Young (1971) reports a clear non-monotonic spacing effect. Relevant results are presented in Fig. 2. Presented here are the data for those trials where there were two presentations (without testing) followed by a test trial (P_1–P_2–T). The retention interval was kept constant at a lag of 10 items while the spacing interval varied from 0 to 17 items. The items consisted of consonant trigrams paired with one of the digits in the range 0–9. We fitted the model to these data and obtained

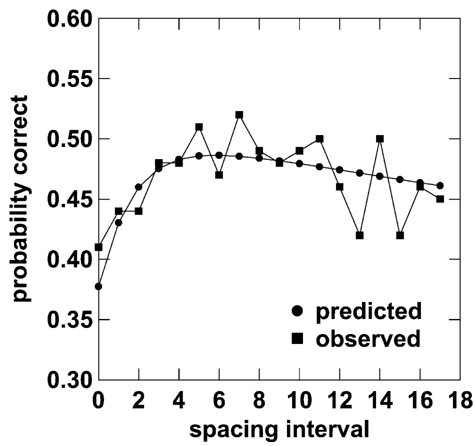

Fig. 2. Observed and predicted probabilities of recall as a function of spacing interval (number of intervening items). Data from Young (1971).

the following parameter estimates: $s = 0.150$, $\lambda = 0.746$, $\alpha = 0.082$, $b = b_2 = 0.246$, $Z = 2.00$, $w = 1.0$ (fixed) and $\theta_2 = 0.30$. The fit of the model is good as indicated by the χ^2-value of 8.47 ($df = 12$) although this is less surprising given the relatively large number of parameters compared to the number of data points. However, the main point of this demonstration is to show that the model is indeed capable of generating non-monotonic spacing effects.

A much more elaborate set of data was collected by Glenberg (1976, Experiment I). In this experiment there were two presentations (without testing) followed by a test trial (P_1–P_2–T), just as in the previous experiment. The paired-associate items were composed of common four-letter nouns constructed so as to avoid obvious pre-experimental associations, rhymes, and orthographic similarities. Fig. 3 gives the observed and predicted probabilities of recall for each of the 24 conditions of the experiment. Each data point is based on 540 observations. The predictions of the SAM model were obtained using the following parameter estimates: $s = 0.260$, $\lambda = 0.800$, $\alpha = 0.013$, $b = b_2 = 0.732$, $Z = 10.0$, $w = 1.0$ (fixed) and $\theta_2 = 0.215$. As can be seen from Fig. 3, the fit of the model is not as good as for the previous experiments, as indicated by the χ^2-value of 41.89 ($df = 18$). However, the model does capture the basic characteristics of the data (except for the very short spacing lags, but see below). It is of some interest to note that this result was already anticipated by Glenberg (1976) who described a semi-quantitative model to account for these data that bears close similarity to the present SAM model.

Although this set of data is frequently cited in discussions of spacing effects, we know of only one other paper that explicitly tried to fit these data. Reed (1977) showed that the Markov model of Young (1971) performed more or less similar as the present model (leading to a χ^2-value of 49.3, $df = 18$). However, a model based on Wickelgren's strength–resistance theory (Reed, 1976; Wickelgren, 1972, 1974a, 1974b) did perform substantially better, leading to χ^2 values between 24 and 30 depending on the exact assumptions of the model. The basic reason for this superior fit seems to be due that the Reed model is capable of fitting the initial

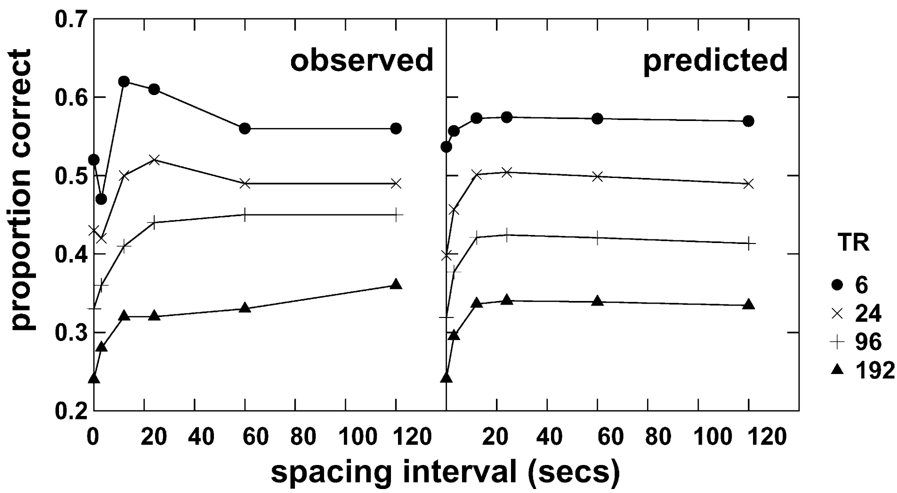

Fig. 3. Observed (left pane) and predicted (right pane) probabilities of recall as a function of spacing interval and retention interval (TR). Data from Glenberg (1976).

decrease for short retention intervals when the spacing interval increases from 0 to 3 s (see the top curve in the left panel of Fig. 3).

However, a close inspection of the model proposed by Reed shows that it contains a rather peculiar assumption in which the items with a lag of 0 are treated quite differently from items that have a lag of 3 s (i.e., one item in between). Items with a lag of 0 are treated as a single presentation and will remain in a short-term buffer for exactly 6 s (just enough to be in the buffer at the time of testing when the retention interval is 6 s, measured from the start of the presentation). However, items with a lag of 3 s will be immediately transferred to a "passive memory" if they are recognized at the second presentation, leading to a somewhat longer functional retention interval and hence a somewhat lower probability of recall (a factor that is compensated at longer retention intervals by the fact that the item starts to build up a "resistance to forgetting" sooner if it is recognized correctly at the second presentation). This strikes us as a rather artificial solution. The solution also depends strongly on the assumption that an item stays in the buffer for exactly 6 s (Reed ran additional fitting runs with this as a free parameter and obtained values in the range of 5.99990–6.00009, which shows that it is indeed crucial that the interval is exactly 6 s). It is hard to defend that an item that is in the buffer at P_2 is treated quite differently from an item that is repeated with a lag of 0.

In order to show that it is this assumption that is responsible for the better fit of the Reed model and not the strength–resistance theory itself, we also ran the SAM model with the assumption that items with a lag of 0 are treated as though they only had a single presentation. For this version, we obtained a χ^2-value of 28.77, comparable to the values obtained by Reed (1977). However, we believe that such an assumption is not really defensible. Moreover, Van Winsum-Westra (1990) performed additional experiments aimed at replicating the dip observed by Glenberg (1976) and could not replicate it. This reinforces our belief that one should not attach too much significance to this small detail. Finally, it should be noted that the Reed model predicts that an item that is presented twice with a single item in between shows worse performance than an item that is presented only once which would be rather astonishing if it were indeed true.

All in all then, the SAM model presented here gives a good account of the major experiments that have varied spacing and retention intervals in a systematic manner. The above applications of the SAM model are not meant as some sort of demonstration that the SAM model is superior to previous models for spacing effects. However, it does show that the results for which in the past dedicated models were developed, can all be fitted within a much more general framework that has also been shown to fit the data from quite different experimental paradigms.

In the next section we will apply this model to a finding that has proved difficult to explain for theories of spacing effects based on variability notions.

6. The Ross–Landauer phenomenon

In 1978, Ross and Landauer published a small paper with rather far-reaching implications. Ross and Landauer took as their starting point the observation that many (if not most) explanations for the typical spacing effect are based on the assumption that spacing effects are due to increasing independence in some relevant attribute as repetitions are more widely separated. Examples of such explanations are the idea that attention fluctuates in some gradual way, the idea that the place where memory traces are stored drifts with time, and the idea that what is stored in the memory traces includes contextual information that fluctuates with time. The present model would appear to fall in the latter category.

Ross and Landauer (1978) argued (and backed this up with mathematical analyses) that any such theory should predict a spacing effect not just for two presentations of a given item but also for two presentations of two different, unrelated items. That is, if there are two unrelated items A and B, and one would look at the probability that *either A or B* is retrieved (the inclusive OR case), then that probability should also show a spacing effect. For example, if the spacing effect depends on the assumption that repetitions that are spaced will lead to the storage of sets of context elements that are more different from each other (due to contextual fluctuation) and that retrieval depends on the overlap of storage and test context, then there should also be a spacing effect for the probability of retrieving either of two different items since these will also share more context elements if they are closer together and hence the probability of recalling *both* items *decreases* as the items are more spaced. That is, $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$ and hence $P(A \text{ or } B)$ will increase if $P(A \text{ and } B)$ decreases, or in other words as the traces become more independent.

At first sight, this reasoning should also apply to the present model. Indeed, Glenberg and Smith (1981, pp. 117–118) argued that this phenomenon provides a problem for the Component-Levels theory of Glenberg (1979), a theory that is quite similar to the present model as we argued earlier. However, as we will show, Glenberg may have been too quick to dismiss his own theory.

A basic assumption of the present model is that information is added to the originally stored trace but only if that trace is retrieved at the second presentation (study-phase retrieval), i.e., if the item is recognized as having been presented before. In order to understand the significance of this assumption for the prediction of the Ross–Landauer phenomenon, let us take the case where the retention interval is relatively long and where the test context, due to random fluctuation, is unrelated to either of the two presentation contexts. For the sake of argument, assume that a repeated item is always recognized at the second presentation. In the present model, the probability of retrieving a trace is directly related to the mean overlap between that trace and the context at test. For a single item presented twice, the overlap between the (combined) trace and the test context will obviously increase as the spacing increases, leading to a higher probability of retrieval at test. However, for the inclusive OR case the situation is quite different. In this case, the test context is (by assumption) not specifically related to either of the two presentation contexts. Hence, the mean overlap with each of the two traces will not vary as the interval between the presentations increases. There may be some residual effect of the correlation between the two presentation contexts but this effect will be much smaller than the effect of the mean overlap in the number of contextual elements.

Thus, if we denote the contextual features by the letters *a*, *b*, *c*, etc. and if we assume that the context at $P_2 = (p, q, r, s)$ and that in the massed case the context at P_1 is equal to P_2 whereas in the spaced case the context is (a, b, c, d) and thus different, then we get for the context stored in each trace:

		MASSED	SPACED
One item twice		(p, q, r, s)	(a, b, c, d, p, q, r, s)
Two items once	I_1	(p, q, r, s)	(a, b, c, d)
	I_2	(p, q, r, s)	(p, q, r, s)

Assume the test context is not specifically related to P_1 or P_2 (i.e., a relatively long retention interval) and is given by (a, c, p, r). In that case the mean overlap with each item is

		MASSED	SPACED
One item twice		(p, r)	(a, c, p, r)
Two items once	I_1	(p, r)	(a, c)
	I_2	(p, r)	(p, r)

Hence, the mean overlap changes for the repeated case but not for the inclusive OR case. Whatever effect of spacing remains, must be due to the change in the correlation between the retrieval of the two traces as a function of the spacing interval and this will be a much weaker effect than that caused by the change in the mean overlap for the repeated case.

In order to verify this reasoning we ran a simulation using exactly the same SAM model as before in which we computed the predicted values for the probability of recalling either *A* or *B* where the spacing interval between *A* and *B* was varied from 1 to 10 and the retention interval was kept constant at 60. In the SAM model as we have applied it here, the contextual fluctuation process is the only factor that produces dependence between the probabilities of recall for two unrelated items as a function of the spacing interval between the two presentations. Since the fluctuation process is random (i.e., any element is just as likely to become active or inactive as any other element), it is possible to compute the probabilities of a specific number of overlapping elements between two moments in time by using standard combinatorial probability theory.

For these computations, we assumed that at any point in time the number of inactive contextual elements was $N = 120$ and the number of active elements was $n = 40$ (hence the total number of contextual elements is set to $N + n = 160$). These values were chosen so as to be similar to the obtained value for s (0.260) in the Glenberg simulation (in the fluctuation model s is equal to $n/(N + n)$). If we define $P_A(P_B)$ as the time at which item $A(B)$ is presented, the model allows us to compute the probabilities for various overlaps in the contextual elements between P_A and P_B as well as between P_B and T. For example, if the number of active elements equals $n = 40$, then the number of elements in the overlap between P_B and T will be binomially distributed with parameters $n = 40$ and p given by Eq. (8) with $A(0) = 1$ and $K = 1$. Similarly, the probability that the overlap between P_A and T equals i elements given that the overlap between P_A and P_B is k and the overlap between P_B and T is j, is then given by

$$P(AT = i \mid AB = k, BT = j) = \sum_{m=0}^{i} \frac{\binom{k}{m}\binom{n-k}{j-m}}{\binom{n}{j}} \times \frac{\binom{n-k}{i-m}\binom{N-n+k}{n-j-i+m}}{\binom{N}{n-j}} \tag{16}$$

That is, the n elements active at P_B are partitioned into a set of k elements that were also active at P_A and a set of $n - k$ elements that were not active at P_A. The j elements in the overlap between P_B and T are drawn randomly from these two subsets, hence the number of elements m that were active at both P_A, P_B, and T follows a hypergeometric distribution (see the first part of Eq. (16)) and similarly for the number of elements $i - m$ that were active at P_A and T but not at P_B (the second part of Eq. (16)). Finally, the overall probability that the overlap between P_A and T is equal to i is given by multiplying these two probabilities (that are independent) and summing over all possible values of m.

Using these equations it is possible to compute the probability of recalling A given recall of B for any spacing and retention interval (since we know the conditional distributions for the contextual overlap and the recall probabilities are a function of this overlap plus constant factors that are determined by the parameters). Fig. 4 gives an example using the parameter estimates obtained for the Glenberg (1976) data. It is clear that the model predicts quite different effects of spacing for the case where a single item is presented twice and the case where two items are each presented once. The two-item-once case shows a gradual decline as a function of the spacing interval (due to more forgetting of A with increasing spacing between A and B) whereas the one-item-twice case shows an increase. In these data, the probability of recalling B given recall of A varied only slightly as a function of the spacing interval: from 0.2502 at a lag of 1 to 0.2500 at a lag of 10 (the unconditional probability was in this case equal to 0.244). Hence, there is some effect due to the contextual fluctuation but the effect is extremely small (note that this analysis disregards dependence due to subject selection effects since we are only interested in changes in the dependence as a function of the spacing interval).

These simulations thus support our intuitive claim that the Ross–Landauer phenomenon is not a problem for the present SAM model for spacing and repetition effects. It follows that to the extent that the Glenberg (1979) theory follows the same principles as the SAM model, these results are also not an embarrassment for Glenberg's Component-Levels theory.

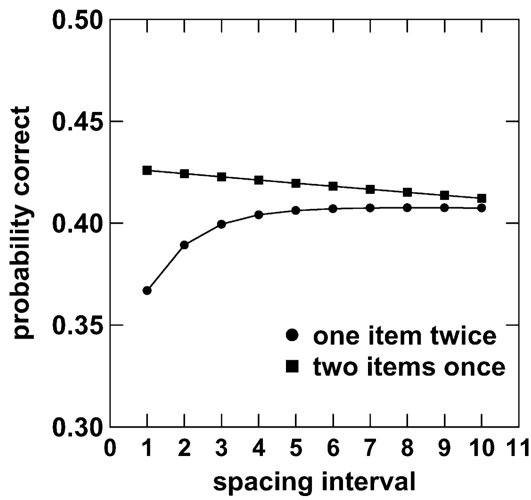

Fig. 4. Predictions of the SAM model for the probability of recalling one or both of two items each presented once, and for the probability of recall of a single item presented twice, as a function of the spacing interval.

7. Concluding remarks

In the present article, we have shown that a fairly straightforward application of the SAM theory coupled with the contextual fluctuation model proposed by Mensink and Raaijmakers (1988, 1989) leads to a viable theory for spacing and repetition effects. In addition to the contextual fluctuation assumptions, a basic principle of the present model is that spacing effects are due to the storage of additional contextual elements in the same trace but only when that trace is retrieved or recognized at the second presentation. Direct support for this assumption is provided by the results of an experiment by Johnston and Uhl (1976). These researchers conducted a test for encoding variability theories of spacing that was based on the idea that if an item is not recognized at the second presentation, this should indicate large differences in encoding (maximal encoding variability) and hence should increase the probability of final recall. Their results showed the opposite: items that were not recognized at the second presentation were recalled poorly. A similar result was obtained by Madigan (1969) who found that a spacing effect was only found for items that were recognized as old at the second presentation. Obviously, there are many problems in interpreting such results (e.g., selection artifacts) but the general finding that spacing effects depend on the recognition or implicit retrieval of the memory trace formed at the first presentation, is clearly consistent with the present model.

We have no intention to provide an extended review of the many and often apparently conflicting results in the literature on spacing effects in human memory nor of the many, mostly qualitative, explanations that have been provided. However, we do believe that the present model and the concepts on which it is based may serve as a general tool for the evaluation of such proposals. For example, some explanations emphasize notions of variability (in some form) while others emphasize deficient processing of massed presentations or the role of study-phase retrieval (see e.g., Greene, 1989). The present model combines a number of such mechanisms (although probably in a somewhat different form) and hence might be used as a vehicle for investigating such less quantitative proposals.

Finally, we would like to draw attention to the fact that the assumption that successive presentations of an item may be stored in the same memory trace, gains indirect support from a quite different application of SAM. Shiffrin, Ratcliff and Clark (1990) examined various models for recognition memory to determine whether or not these models could be made consistent with the so-called *list-strength effect* (see Ratcliff, Clark, & Shiffrin, 1990). A modified version of the SAM model for recognition (Gillund & Shiffrin, 1984) could be made to predict such a result (using the *differentiation assumption*, see Shiffrin et al., 1990) but only if it was assumed that repeated presentations of the same item were encoded in a single memory trace. Additional evidence for this explanation was provided by Murnane and Shiffrin (1991), see also Shiffrin, Murnane, Gronlund, and Roth (1989). They tested whether a reversal of the list-strength effect in recognition occurs if repetitions are presented in such a way that they are likely to be encoded in separate images. They found that repetitions of words in different sentences produced a list-strength effect whereas repetitions of entire sentences did not. As a direct corollary, the present model for spacing effects would predict that in such a paradigm in which repetitions are not encoded in a single trace, the effects of spacing between repetitions would be greatly diminished.

Even though the present model seems to provide an adequate framework for the analysis of spacing effects, there are still some remaining problems. First of all, the model in its full form contains a relatively large number of parameters. For many experimental results several of the parameters may be fixed at some relatively arbitrary value without any noticeable effect on the goodness-of-fit of the model. Thus, in many cases the parameters of the full model will not all be identifiable making it very hard to meaningfully compare the parameter values between experiments. In addition, we have found that many variants of the model in which slightly different assumptions are made (e.g., with respect to what happens with items in the buffer on the second presentation) will produce more or less similar results. Thus, it appears that the data of these experiments do not completely determine the exact form of the model. However, we have not been able to come up with a completely satisfactory rationale that would clearly favor one version over another on theoretical or empirical grounds. In conclusion then, despite the fact that the data that we have fitted represent the best data sets available (being based on relatively large numbers of observations and varying relevant dimensions such as spacing and retention intervals in a parametric manner), many of the nitty-gritty details can only be defended on such grounds as simplicity and elegance.

As mentioned in the introduction, the present application of SAM in combination with the results presented in previous papers shows that the SAM theory provides a general framework that can successfully explain a large number of the basic findings in a variety of episodic memory paradigms. In that sense, the SAM project has more than fulfilled its initial aims and hopes. We are now in a position to see whether it is possible to extend the framework to paradigms from other areas of human memory, in particular semantic memory and implicit memory phenomena. In recent years, we have been working on various projects to develop models for priming effects in perceptual identification and lexical decision based on the REM framework (see e.g., Raaijmakers & Shiffrin, 2002; Schooler et al., 2001; Wagenmakers et al., 2001). As mentioned before, the REM model is closely similar to SAM (especially with respect to recall paradigms). It is of some interest to note that the REM approach to semantic memory assumes that semantic memory arises out of episodic memory traces due to the adding of information to previously stored traces. Thus, the present assumption regarding the importance of study-phase retrieval effects may have a much broader significance than the prediction of spacing effects alone.

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. 2, pp. 89–195). New York: Academic Press.

Baddeley, A. D., & Longman, D. J. A. (1978). The influence of length and frequency of training sessions on the rate of learning to type. *Ergonomics, 21*, 627–635.

Bahrick, H. P., & Phelps, E. (1987). Retention of Spanish vocabulary over 8 years. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 344–349.

Bower, G. H. (1961). Application of a model to paired-associate learning. *Psychometrika*, *26*, 255–280.

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

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

Estes, W. K. (1960). Learning theory and the new "mental chemistry". *Psychological Review*, *67*, 207–223.

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

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

Glenberg, A. M. (1979). Component-Levels theory of the effects of spacing of repetitions on recall and recognition. *Memory* & *Cognition*, *7*, 95–112.

Glenberg, A. M., & Smith, S. M. (1981). Spacing repetitions and solving problems are not the same. *Journal of Verbal Learning and Verbal Behavior*, *20*, 110–119.

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.

Gronlund, S. D., & Shiffrin, R. M. (1986). Retrieval strategies in recall of natural categories and categorized lists. *Journal of Experimental Psychology: Learning Memory and Cognition*, *12*, 550–561.

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.

Hintzman, D. L. (1976). Repetition and memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 10). New York: Academic Press.

Howard, M. W., & Kahana, M. J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning Memory and Cognition*, *25*, 923–941.

James, F., & Roos, M. (1975). MINUIT, a system for function minimization and analysis of the parameter errors and correlations. *Computer Physics Communications*, *10*, 343–367.

Johnston, W. A., & Uhl, C. N. (1976). The contribution of encoding effort and variability to the spacing effect on free recall. *Journal of Experimental Psychology: Human Learning and Memory*, *2*, 153–160.

Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory* & *Cognition*, *24*, 103–109.

Landauer, T. K. (1967). Interval between item repetition and free recall memory. *Psychonomic Science*, *8*, 439–440.

Landauer, T. K. (1969). Reinforcement as consolidation. *Psychological Review*, *76*, 82–96.

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

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. (1970). The situation with respect to spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behavior*, *9*, 596–606.

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

Mensink, G. J. M., & Raaijmakers, J. G. W. (1989). A model of contextual fluctuation. *Journal of Mathematical Psychology*, *33*, 172–186.

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

Raaijmakers, J. G. W. (1979). *Retrieval from long-term store: A general theory and mathematical models*. Doctoral dissertation, University of Nijmegen, The Netherlands.

Raaijmakers, J. G. W., & Phaf, R. H. (1999). Part-list cuing revisited: A test of the SAM explanation. In C. Izawa (Ed.), *On memory: Evolution, progress and reflection on the 30th anniversary of the Atkinson–Shiffrin model* (pp. 87–104). Mahwah, NJ: Lawrence Erlbaum.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 14, pp. 207–262). New York: Academic Press.

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

Raaijmakers, J. G. W., & Shiffrin, R. M. (2002). Models of memory. In H. Pashler & D. Medin (Eds.), *Stevens' handbook of experimental psychology: Memory and cognitive processes* (3rd ed., Vol. 2, pp. 43–76). New York: Wiley.

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

Reed, A. V. (1976). List length and the time course of recognition in immediate memory. *Memory* & *Cognition*, *4*, 16–30.

Reed, A. V. (1977). Quantitative prediction of spacing effects in learning. *Journal of Verbal Learning and Verbal Behavior*, *16*, 693–698.

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.

Rumelhart, D. E. (1967). *The effects of interpresentation intervals on performance in a continuous paired-associate task*. Technical report 16. Institute for Mathematical Studies in Social Sciences, Stanford University.

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

Shiffrin, R. M. (1970). Memory search. In D. A. Norman (Ed.), *Models of human memory* (pp. 375–447). New York: Academic Press.

Shiffrin, R. M., Murnane, K., Gronlund, S., & Roth, M. (1989). On units of storage and retrieval. In C. Izawa (Ed.), *Current issues in cognitive processes: The Tulane Flowerree symposium on cognition* (pp. 25–68). Hillsdale, NJ: Lawrence Erlbaum.

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

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

Smith, S. M., & Rothkopf, E. Z. (1984). Contextual enrichment and distribution of practice in the classroom. *Cognition and Instruction*, *1*, 341–358.

Van Winsum-Westra, M. (1990). *Spacing and repetition effects in human memory*. Doctoral dissertation, University of Nijmegen, The Netherlands.

Wagenmakers, E. J. M., Steyvers, M., Raaijmakers, J. G. W., Shiffrin, R. M., van Rijn, H., & Zeelenberg, R. (2001). *A Bayesian model for lexical decision*. Submitted for publication.

Wickelgren, W. A. (1972). Trace resistance and decay of long-term memory. *Journal of Mathematical Psychology*, *9*, 418–455.

Wickelgren, W. A. (1974a). Single-trace fragility theory of memory dynamics. *Memory* & *Cognition*, *2*, 775–780.

Wickelgren, W. A. (1974b). Strength/resistance theory of the dynamics of memory storage. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology*. New York: Freeman.

Young, J. L. (1971). Reinforcement-test intervals in paired-associate learning. *Journal of Mathematical Psychology*, *8*, 58–81.