Raaijmakers, J.G.W & Phaf, R.H. (1999). Part-list cuing revisited: Testing the sampling-bias hypothesis. In C. Izawa (Ed.), *On memory: Evolution, progress and reflection on the 30th anniversary of the Atkinson-Shiffrin model*. (Pp 87-104). Mahwah, N.J.: Lawrence Erlbaum Associates.

Chapter 5

Part-List Cuing Revisited: Testing the Sampling-Bias Hypothesis

Jeroen G.W. Raaijmakers
R. Hans Phaf
University of Amsterdam, the Netherlands

The SAM (Search of Associative Memory) model was originally developed as a model for free and cued recall during the beginning of 1978. After the initial simulations of a number of standard findings in free recall had shown that the model performed reasonably well (see Raaijmakers & Shiffrin, 1980), the model was applied to the part-list cuing effect, an intriguing finding that was a real puzzle at that time (and perhaps still is, see Nickerson, 1984). This effect refers to the phenomenon in a free recall paradigm that presenting some of the list items as cues for recall does not help subjects in improving their retrieval of the remaining items (the "targets"). This paradigm was devised by Slamecka (1968) as a test of the general assumption that memory is associative and that such associations should aid the retrieval of associated items from memory. Thus, it was assumed that the retrieval of the list items from memory should be facilitated when subjects are given some of them as cues. Hence, the finding that no positive effect was obtained and that often the cues even seemed to have a slight effect was a real puzzle.

In May 1978, the first simulations of this paradigm were run with the SAM model. The results showed that the model predicted the part-list cuing effect, even though the assumptions of the model were such that it made heavy use of interitem associations. Although the model predicted this result, we did not understand very well why. We initially thought that this prediction resulted from the assumption in the model that associations were strengthened (“incremented”) following successful recall. However, in November 1978, in a class project at Indiana University, one of Shiffrin’s students ran a simulation of the model with all strengthening parameters set to zero. To our surprise, the effect did not disappear.

Given the structure of the SAM model and the observation that the effect (at least within the model) did not depend on the incrementing assumptions, it had to be the case that the prediction was due to the fact that the cues used by the noncued subjects (the cues they generated themselves during the recall process) were somehow better than the randomly selected part-list cues that were used by the cued subjects. One reason for the superiority of self-generated cues might be that these were not sampled randomly but with a probability that was related to their strength. Thus, these items presumably were stronger and hence might also have stronger interitem associations than the experimenter-provided part-list cues.

In order to understand more fully what was going on, we ran a number of simulations with a very simplified version of SAM (Raaijmakers, 1979). In this version, all items were of equal strength and the only process that played a role was the sampling process (i.e., "recovery" was always successful). A very simple associative structure was assumed in which each item was associated to exactly one other item. In these simulations a negative part-list cuing effect was obtained when the cues were drawn at random from all list items, even though the setup was such that strong positive effects would have been obtained if the cues had been selected in concordance with the stored associative structure, that is, one cue per pair (as in regular paired-associate recall). These analyses showed that there was a subtle factor that favored the noncued group. It turned out that the cues used by the cued group were inferior to the self-generated cues used by the noncued group due to a sampling bias: the self-generated cues could provide access to the most "profitable" clusters of interassociated items, those composed of all target items, and the experimenter-provided cue items could not do this (since these were by definition not associated to the items in those clusters).

This explanation was proposed by Raaijmakers and Shiffrin (1981a) as the major explanation for why the SAM model predicted the part-list cuing effect. They (p. 114) pointed out that the prediction was crucially dependent on the fact that the associative strength distribution was not uniform since the effect becomes positive when all strengths are set to the same value (hence the normally existing structure in the matrix of retrieval strengths is eliminated). Additional analyses gave further support for the SAM explanation. In particular, it was shown that the model predicted a positive effect of cuing when the context-to-item associations were very weak. In such a case, the noncued group would no longer be able to self-generate enough cues: Bad cues are still better than no cues at all. Experiments by Blake and Okada (1973) and Basden (1973) indeed showed positive cuing effects when the cues were given following interpolated study of an interfering list.

Over the years, several other explanations have been provided for the part-list cuing effect. Most of these explanations assume (in one way or another) that subjects do not or cannot make use of interitem associations to aid recall. Other explanations assume that the normal retrieval process in the cued condition is somehow disrupted and hence less effective than that in the control condition. For example, Basden, Basden and Galloway (1977) proposed that the effect is due to the fact that the order of recall in the cued group is suboptimal because it is not compatible with the organization that was stored during study of the list. However, the explanation proposed by Raaijmakers and Shiffrin (1981a) still seems the only one that is capable of explaining not just the basic result, but also a range of variations, including those situations in which a positive effect is obtained. Most importantly, the SAM explanation does not rely on the awkward assumption that in free recall no use is made of interitem associations. On the contrary, the sampling bias explanation is critically dependent on the fact that such associations are effective.

Not everyone was, however, convinced by our analyses. For example, Nickerson (1984) in his review of the literature, criticized our explanation because it made "some assumptions the only justification for which seems to be that they are needed to explain the effect" (p. 550). He was referring to the assumption that both groups sample the same number of clusters. Although he did acknowledge that this was only assumed for the simplified illustration in the Raaijmakers and Shiffrin (1981a) article, Nickerson completely disregarded the fact that in the full SAM model this was not in fact assumed a priori. The simplified illustration served only to explain a rather subtle factor that had been missed in all previous discussions of the part-list cuing phenomenon (i.e., the sampling bias factor). The crucial aspect was not that an equal number of clusters was accessed but that the cue items that were used by the cued group could only give access to clusters of items that had at least one cue in them. The "assumption" that the cued group would access about the same number of clusters was not an a priori assumption, but rather could be derived from the general retrieval assumptions (i.e., the assumption that the efficiency of retrieval decreases with each additional retrieval). Hence, Nickerson's objection seems unfounded.

A more crucial objection was advanced by Roediger and Neely (1982). They described an experiment by Park (1980) that showed that when words were embedded in sentences, part-list cues did aid recall. However, when the same items were embedded in categorized lists, the usual negative part-list cuing effect was observed. A similar reversal was observed when subjects studied sets of word triples either by forming coherent interactive images or separate images. Roediger and Neely proposed a distinction between horizontal (direct interitem) and vertical associations. According to them, only when there are horizontal associations between items (direct item-to-item associations), part-list cuing leads to facilitation. When there are no horizontal associations, the result is inhibition. They pointed out that this result is difficult to reconcile with the SAM explanation since SAM already uses a rich interitem associative structure in its prediction of part-list cuing inhibition.

In addition, the SAM model has been criticized as being overly complicated and of using a large number of parameters (see Roediger, 1993), hence making it easy to predict any result. Moreover, "designing independent tests of the model may be difficult, since its ten parameters and numerous countervailing processes make unambiguous predictions hard to come by" (Roediger & Neely, 1982, p. 225). Although this objection seems unfair when a quantitative model like SAM is compared to verbal, descriptive theories (a nonquantitative version of SAM would surely be more powerful in the sense of being consistent with an ever larger set of data patterns since it is less constrained), it does point to the importance of being able to devise proper experimental tests of the model. In this chapter, we will present a series of experiments that were motivated by the SAM explanation for the part-list cuing effect and that may provide a critical test of our explanation against explanations such as the one proposed by Roediger and Neely (1982).

In particular, we wanted to show that even in cases where the normal testing procedure leads to part-list inhibition, a positive effect could be obtained using a different testing procedure. Basically, the idea was to give the recall test either immediately after study or after a delay filled with the learning of an unrelated list. The latter should lead to retroactive interference and hence recall levels should be much lower after the interpolated learning. In SAM, this would be equivalent to the assumption that the context strengths were much lower in the delayed testing case than in immediate testing (see Mensink & Raaijmakers, 1988). As was mentioned above, Raaijmakers and Shiffrin (1981a) showed that SAM predicts a positive part-list cuing effect if the context strengths are low, since in that case the noncued group will not be able to generate a sufficient number of items to use as cues whereas the cued group still has the experimenter-provided cues to use in their retrieval.

A second crucial aspect of the SAM explanation was that the cuing effect would become positive if the cues are not sampled randomly (as is normally the case) but are chosen in such a way that they would give maximum access to the stored clusters. For example, the effect should become positive if the list consists of a number of categories and subjects are given one cue per category. Thus we assumed that if one would let subjects study a list of paired associates and then at test would give half of the items as cues, the part-list cuing effect should be negative if the cue words are sampled randomly and positive if one cue word is chosen from each pair of words. The simplified model that had been used to analyze SAM's prediction of the part-list cuing effect indeed predicted such a result.

In the first experiment, we presented subjects lists of paired associates. There were five test conditions, three immediate and two delayed ones. The three immediate cuing conditions were: no cues (the *control* condition), random half of the words as cues (the *random* condition), or one cue from each pair (the *1Q* condition). The two delayed conditions were: no cues or a random half of the words. Based on the analyses described above, we expected that there would be a negative part-list cuing effect in the immediate condition for randomly chosen cue words but a positive effect for the condition with one cue per pair. On the delayed condition, the part-list cuing effect should reverse, hence the random cue condition should now be superior to the no cue control condition.

Experiment 1

Method

Subjects. Thirty-two volunteer subjects participated in Experiment 1. All subjects were undergraduate psychology students at the University of Nijmegen who participated for course credit. Subjects were tested in groups of 2-5 subjects at a time.

Materials and Design. There were five conditions in the experiment. In all conditions, subjects were asked to study a list of 20 unrelated word pairs (hence 40 words in total). In three conditions, recall was tested immediately, in two conditions recall was tested after a delay filled with three study-test cycles of an unrelated list of 25 word pairs. In the immediate testing conditions, either no cues were given (*control*), or 20 randomly chosen words were given as cues (*random*), or one cue from each pair was given (either the first or the second, *1Q*). In delayed testing, either no cues were given or 20 randomly chosen words were given as cues. All words were common Dutch nouns. Each subject participated in all three immediate testing conditions and in one of the two delayed test conditions. As each subject participated in four conditions, four lists of 20 unrelated word pairs were constructed. The order of the conditions and the assignment of lists to conditions were counterbalanced across subjects.

Procedure. Subjects were seated in front of a monitor. The subjects booths were separated by screens. After a practice list, the four experimental lists plus the three repetitions of the interpolated list were presented. The word pairs were presented for 4 seconds each. After each presentation of a list, a three digit number was presented on the screen followed by 10 single-digit numbers that had to be subtracted from the three-digit number. Each number was presented for 2 seconds, hence this task (that was used to eliminate recall from STS) took about 20 seconds. At the end, they wrote down the answer on their answer sheet.

After the arithmetic task, subjects were either tested immediately or they were given the interpolated learning task. At test, they were asked to recall in writing (answer sheets were provided) as many words as possible from the list presented last, in any order. They were given two minutes to recall. After each

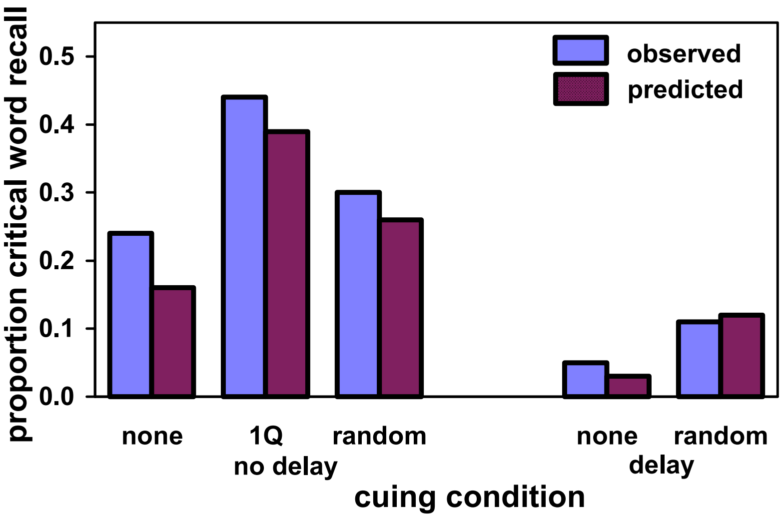

FIG. 5.1: Observed and predicted proportions of critical word recall for each cuing condition in Experiment 1. Predictions were obtained from the SAM model (adapted to study of word pairs) with r = 2, a = 0.3, b = 0.8, c = 0.3, d = 0.02, e = f = g = 2, K_{MAX} = 50 and L_{MAX} = 2. For delayed testing, the context strength was decreased to a = 0.05.

test, the answer sheets were collected by the experimenter. The instructions explained that on some tests a number of words from the list would be presented on the screen and that these could be used to recall other words that they had been associated with. The interpolated list consisted of 25 unrelated word pairs. Each pair was presented during study for 3 seconds. After all pairs had been presented, a test was given of the interpolated list in which the first member of the pair was given for subjects to write down the second member. Response time was six seconds per tested item. The whole session lasted about one hour.

Results and Discussion

In Fig. 5.1 the proportions of noncue words (targets) recalled in each condition are shown. Two repeated-measures analyses of variance were performed on these data, one for the immediate testing conditions and one for the delayed testing conditions. All differences between conditions were significant at α =

0.05. As expected, in immediate testing performance in the 1Q condition was superior to the other two conditions. However, the random cue condition was also superior to the no-cue control condition. We had expected to find a standard negative part-list cuing effect based on the simulation results with the simplified version of SAM. In delayed testing, the cuing effect was positive, as expected. However, since the effect was also positive in immediate testing, we obviously did not obtain the reversal of the effect as we had expected.

These results replicate the positive cuing effects obtained by Park (1980, see Roediger & Neely, 1982) when during study subsets of words are strongly interassociated, such as by forming interactive images. These results seem to be inconsistent with the explanation given by SAM for the part-list cuing effect since SAM has been shown to predict a negative effect in free recall even when the interitem associations are strong. However, these simulation results were obtained with the normal simulation program for free recall in which items are presented one at a time. In such a case, the clusters formed during study will be overlapping. In the present case, however, the items are presented in pairs. If it is assumed that interitem associations will only be formed between the two members of a pair, the resulting clusters will be non-overlapping. In previous analyses of SAM (see Raaijmakers & Shiffrin, 1981b) we did indeed assume that when paired-associates are studied, the buffer would contain only the two members of the pair being studied and hence words belonging to different pairs would not become associated. On the other hand, simulations with the simplified model had led us to expect a negative part-list cuing effect even for lists of pairwise associated items.

The results of this experiment prompted us to take a closer look at the predictions of the full SAM model, adapted to this experimental design. We ran a number of simulations, assuming that items were studied in pairs. As in Raaijmakers and Shiffrin (1981b), the amount of information stored on a trial is assumed to be a function of the number of items in the buffer. Thus, if there are 2 items in the buffer, the increase in contextual associative strength will be $a/2$ for each second of study. Otherwise, the simulation program was exactly the same as that used in the original analyses of the part-list cuing paradigm.

To our surprise, the results of these simulations showed that the SAM model did in fact predict a positive cuing effect when the items are studied in pairs. In Fig. 5.1 the results are shown based on 500 simulation runs with the following parameter values (no attempt was made to search the parameter space exhaustively in order to optimize the fit of the predictions): $r = 2$, $a = 0.3$ (immediate testing) or $a = 0.05$ (delayed testing), $b = 0.75$, $c = 0.25$, $d = 0.02$, $e = f = g = 2$, $K_{MAX} = 50$ and $L_{MAX} = 2$ (see Raaijmakers & Shiffrin, 1981a, p. 97, for an explanation of the parameters of the SAM model). Hence, it was assumed that delayed testing leads to a decrease in the context-to-image strengths but leaves all other aspects unchanged. The SAM model predicts all of the effects that were obtained in this experiment: It shows a huge advantage for the 1Q condition and a slightly superior performance in the random cue condition compared to the no cue control condition, both in immediate and delayed testing.

Hence, despite the fact that the experiment was in some sense not successful because it did not produce the expected reversal of the cuing effect, it did lead to the surprising discovery that the SAM model predicts positive cuing effects if the list consists of a number of small and non-overlapping clusters. The reader will probably wonder (just as we did) why the simplified model did not predict this. Further analyses showed that this was due to the fact that in the simplified model the recovery process was eliminated. If in the full model one sets the probability of successful recovery to 1 (irrespective of the cues used), the predicted part-list cuing effect will always be negative. This is probably due to the fact that such an assumption eliminates one of the few factors in the model that favors the cued condition, relative to the control condition. In the control condition relatively more searches are made using only the context cue, whereas in the cued condition almost all searches are made using both context and item cues. In the latter case, the probability of successful recovery will be greater than when only the context cue is used.

In sum, we have shown that, contrary to what we and others had originally assumed, the full SAM model predicts a positive cuing effect if the list consists of a number of highly interassociated and non-overlapping clusters of items. This is, of course, a highly interesting result, but it also implies that the present experimental setup will not make it possible to test the SAM model against the Roediger-Neely hypothesis that assumes a direct relationship between the stored associative structure and the direction of the part-list effect. In order to decide between these two theories, we need a design in which the part-list cuing effect is negative in immediate testing. Therefore, in the next experiment we gave subjects regular free recall lists (lists of words presented one at a time). Otherwise, the experimental procedures were the same as the previous ones. Subjects were tested either immediately or after a delay filled with study of an unrelated list of paired-associates. One obvious problem with the normal free recall procedure is that one has no control over the clusters that will be formed by a subjects. Hence, it will not be possible to give subjects one cue from each cluster at test. However, on the assumption that consecutive items might be more likely to become interassociated, we included a condition that resembled the one cue per cluster condition of Experiment 1. In this case, we gave subjects either all the items in the even positions as cues or all the uneven items. Obviously, there was no guarantee that this would coincide with the stored associative structure.

Experiment 2

Method

Subjects. Thirty paid subjects from the subject pool of the TNO Research Institute for Human Factors participated in the experiment. Subjects were tested individually.

Materials and Design. There were six conditions in the experiment. In all conditions, subjects were asked to study a list of 40 words. In three conditions, recall was tested immediately, in the other three conditions recall was tested after a delay filled with three study-test cycles of an unrelated list of 25 word pairs (which lasted about 15 min). In both immediate and delayed testing, either no cues were given (*control*), or 20 randomly chosen words were given as cues (*random*), or the 20 words that had been presented in either the even serial positions or the uneven positions (we will still refer to this as the *1Q* condition). All words were common Dutch nouns. Each subject participated in all conditions. Six lists of 40 words were constructed, and the assignment of lists to conditions as well as the order of the conditions were counterbalanced across subjects.

Procedure. The procedure was similar to that of Experiment 1, except that now the words were presented one at a time, for 2 seconds each. The whole session lasted about two hours. There was a break about halfway during the session.

Results and Discussion

For each subject, the number of target words recalled was determined for each condition with cues. Two scores were computed for each of the two conditions without cues, one score as a control for the random cues condition and one as a control for the 1Q condition (two scores are needed since the target items are different depending on the cues that are actually given). A 6 x 2 x 2 x 2 repeated-measures analysis of variance with groups as between-subjects factor and delay, cue type (random vs. 1Q) and cuing (with or without cues) was performed. As expected, there was a large effect of delay: $F(1, 24) = 110.9$, $p < .001$. None of the other main effects was significant. Thus, the cue type factor neither generated any effect nor interacted with any of the other factors. Hence, there was no advantage for the 1Q condition relative to the random cue condition. Apparently, the associative structure cannot be characterized as one of pairwise associations between consecutive items. There was, however, a significant interaction between delay and cuing: $F(1, 24) = 10.8$, $p < .005$. Figure 5.2 shows the nature of this interaction: the effect of cuing is negative in immediate testing, but positive in delayed testing. We have, here, confirmed the

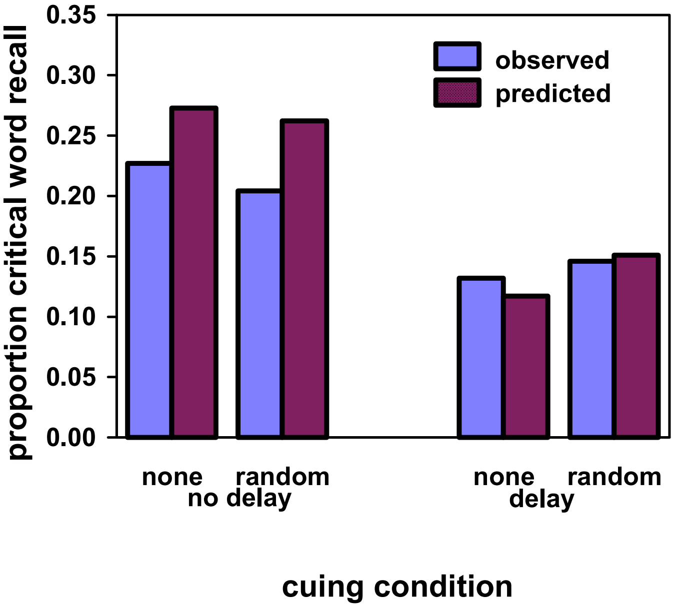

FIG. 5.2: Observed and predicted proportions of critical word recall for the cued and noncued conditions in Experiment 2. Predictions were obtained from the SAM model with r = 4, a = 0.3, b = 0.1, c = 0.1, d = 0.02, $e = f = g = 2$, K_{MAX} = 50 and L_{MAX} = 2. For delayed testing, the context strength was decreased to a = 0.10.

prediction of SAM that the part-list cuing effect reverses when unaided retrieval (as in the control condition) becomes difficult.

In Fig. 5.2 we also present the predictions from SAM. These predictions are based on 500 simulation runs with the same model as was used to generate the predictions for Experiment 1, but now adapted to regular free recall. Thus, the traditional buffer model was used, but the amount of information stored on a trial is a function of the number of items in the buffer. The following parameter values were used: r = 4, a = 0.3 (immediate testing) or a = 0.10 (delayed testing), b = 0.10, c = 0.10, d = 0.02, $e = f = g = 2$, K_{MAX} = 50 and L_{MAX} = 2. Again, these predictions are representative and are not based on any elaborate search of the parameter space. These simulation results demonstrate that the SAM model indeed predicts the observed reversal of the cuing effect in delayed testing.

These results, therefore, confirmed the prediction of the SAM model that the part-list cuing effect might reverse if the context strength is low. The Roediger-Neely hypothesis is not consistent with these results, because in their analysis the direction of the part-list cuing effect is determined by the presence or absence of horizontal associations. In such a model, the effect depends on how the information was stored during study and should not depend on whether it is tested immediately or after a delay.

Experiment 3

In Experiment 3, we further analyzed the basic prediction of SAM that the part-list cuing effect critically depends on the relation between the stored associative structure and the cues provided at test. To this end, we must be able to manipulate the stored associative structure. We, therefore, gave subjects categorized lists, i.e., lists in which several words belonged to the same category. On the assumption that the stored associative structure reflects to some extent this predefined categorical structure, it should be possible to manipulate the relation between the cues provided at test and the stored associative structure.

Method

Subjects. Twenty-five subjects participated in the experiment. Eight of them were undergraduate students who took part in the experiment for course credit. The other subjects, also mainly students, were paid volunteers. Seven experimental sessions were run, each with either 3 or 4 subjects.

Materials and Design. Fifty-one categories were taken from Dutch category norms (Hudson, 1982). Four words were selected from each category. The most frequently produced word was left out because this word has a high probability of being produced solely due to the pre-experimental association. Five lists of 10 categories each were constructed in such a way that no list had any related categories. In addition, a practice list was constructed containing only five categories, four of which were not taken from any category norms, but were made up by the experimenter.

Recall of all five experimental lists was tested in one session. The tests differed in the way the cues were chosen from the list. In the first condition no cues were given (control condition). In the second condition 10 cues were given, one from each category (1Q condition). The other conditions also had 10 cues, but these were either chosen randomly (random condition, condition 3) or in such a way that there were five categories with 2 cues each, and 5 categories with no cues. In condition 4, the cued categories were blocked, that is, categories 3-7 of the presentation order were cued, or distributed over the list, that is, all even-numbered categories were cued. These last two conditions will be referred to as the blocked and distributed conditions. In order to avoid confounding due to primacy effects, the first presented category was never cued.

The order of the five conditions as well as the assignment of lists to each condition was balanced over subjects. The practice list was the same for all subjects.

Procedure. The procedure was similar to those of Experiments 1 and 2. Lists were presented in blocks of four words. A block consisted of all members from one category and was presented for five seconds. During the test, the cues (if any) were shown on the screen and subjects were instructed to read the cue words before writing down any responses. Subjects had three minutes to recall. The cues remained on the screen during the entire recall period. The whole session took about 50 minutes.

Results and Discussion

As shown in Fig. 5.3, the proportion of noncue words (targets) recalled from a categorized list varied considerably depending on the way in which the cues have been selected from that list. An analysis of variance on the number of recalled words showed a highly significant effect for the conditions: $F(4, 96) = 9.89$, $p < .001$. A posteriori tests (Tukey) showed that this was mainly due to the high performance in the 1Q condition: $q(96) = 6.12$, $p < .01$. Interestingly, as in Experiment 1, the effect of cuing in the random cue condition was not negative, but slightly positive (although this was not statistically significant). There was also a significant difference between the random cue condition and the 2Q-distributed condition (2 cues for 5 categories, distributed over the list): $q(96) = 4.04$, $p < .01$.

The nature of these effects can be seen more clearly if we examine the

Table 5.1
Proportion of Critical Words Recalled for Cued and Noncued Categories Separately in Experiment 3.

	no cues	*1Q*	*random*	*2Q block.*	*2Q distr.*
from cued categories	--	0.627	0.648	0.727	0.732
from noncued categories	0.455	--	0.195	0.306	0.225

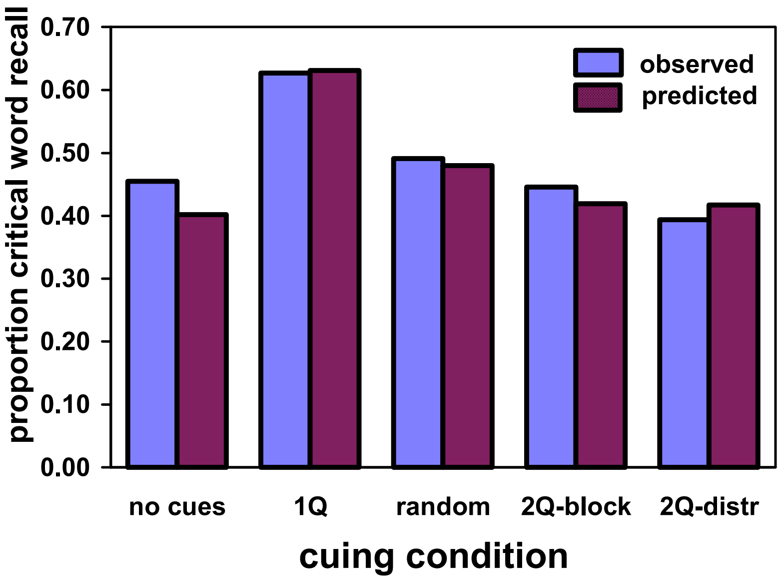

FIG. 5.3: Observed and predicted proportions of critical word recall for the cued and noncued conditions in Experiment 3. Predictions were obtained from the SAM model for categorized lists with *r* = 4, *PCS* = 1.30, *a* = 0.1, *b* = 0.0, *c* = 0.1, *d* = 0.02, *e* = *f* = *g* = .7, K_{MAX} = 50 and L_{MAX} = 3 (see text for explanation).

number of words recalled from cued and noncued categories separately (see Table 5.1). Whereas cuing has a positive effect on the recall performance for items belonging to the cued categories, the effect on recall of the members of the noncued categories is clearly negative. The proportion items recalled from non-cued categories declines with the number of items recalled from the cued categories. Since recall from cued categories precedes recall from noncued categories, this may be interpreted as an indication that recall of an item reduces the probability of recall of a further item.

This is further illustrated by the data in Table 5.2 in which the proportion of categories from which at least one item is retrieved and the proportion of critical items recalled from these 'retrieved' categories are given. It is evident that the probability of retrieving a category increases with the number of cues provided from that category. However, the probability of retrieving a noncued

Table 5.2
Proportion of Categories Retrieved and Proportion of Critical Words Recalled from Retrieved Categories for Cued and Noncued Categories Separately in Experiment 3.

		no cues	*1Q*	*random*	*2Q block.*	*2Q distr.*
categories retrieved	*cued*	--	0.83	0.80	0.87	0.89
	noncued	0.53	--	0.25	0.39	0.30
critical items recalled	*cued*	--	0.76	0.81	0.83	0.82
	noncued	0.86	--	0.78	0.78	0.76

category decreases with the number of retrieved cued categories. When no categories are cued, 53% of the noncued categories are retrieved, while in the random cue condition (in which on the average 7.4 categories were cued), only 25% of the noncued categories were retrieved. However, once a category is retrieved, the proportion of recalled critical items from that category is approximately constant. Such effects can be explained by the assumption that as more and more categories are retrieved, there is increasing output interference, making it more and more difficult to retrieve other categories (or items from other categories).

Thus, in the present experiment both positive and negative cuing effects were observed. These results support the hypothesis that the critical factor in the sign of the part-list cuing effect is not the use of horizontal or vertical associations but the selection of the cues relative to the stored associative structure.

We also analyzed whether the SAM model could account for these effects. In order to maximize the contrast of the SAM model with theories assuming vertical associations, it was assumed that category information was completely captured by the interitem associations between items belonging to the same category. Thus, all associations between items belonging to the same category were given a relatively high pre-experimental associative strength (*PCS*, pre-experimental category strength). The association strengths resulting from the study of the items were added to these pre-experimental strengths. However, for the present experiment the pre-experimental strengths are indistinguishable from the experimental interitem strengths because all members of a category were always studied together. In the simulations we have, therefore, set PCS to a relatively high value and the normal interitem strength parameter (*b*) to zero. Otherwise, the standard SAM assumptions and parameter definitions were used (e.g., *a* gives the amount of contextual strength added per second of study). Figure 5.3 gives representative predictions based on 100 simulation runs with the parameter values $r = 4$, $PCS = 1.30$, $a = 0.1$, $b = 0.0$, $c = 0.1$, $d = 0.02$, $e = f = g = .7$, $K_{MAX} = 50$ and $L_{MAX} = 3$. SAM correctly predicts strong positive effects of cuing with one cue per category, either small negative or small positive effects in the random cuing condition and larger negative effects in the two cues per category conditions. The simulation model that was used has no mechanism to predict a difference depending on how the categories from which the two cues are given, are distributed across the list. In order to explain such a difference, the model had to be adjusted by introducing different contexts for different parts of the list (i.e., different contexts for the beginning, middle, and end of the list). Simulations with such an adjusted model indeed produced the observed pattern. However, we will not pursue this second-order effect further since it is not important for the central issue of this chapter, namely the issue of whether the part-list cuing effect can be explained by a model that relies on horizontal associations.

General Discussion

One of the most intriguing findings obtained in the present experiments was the unexpected (small) positive effect of cuing when the study list consists of either paired-associates or a number of small-sized categories. In the past, positive part-list cuing effects have been interpreted as being due to the formation of horizontal associations during study of the list (Roediger & Neely, 1982). The direction and size of the effect, however, were shown here to depend critically on two factors that only affect the retrieval process rather than the stored associative structure. In Experiments 1 and 3, the effect depended, to a large degree, on whether the cues were chosen randomly or in such a way that they were distributed evenly across the associative clusters formed during study. In Experiment 3, moreover, there was a strong positive cuing effect if one cue was chosen from each category but the effect was negative if two cues were chosen from each of five categories and no cues from the remaining five categories.

In addition, we experimentally verified the prediction of SAM that the direction of the part-list cuing effect may reverse if it becomes difficult to retrieve items without any experimenter-provided cues. In Experiment 2, the effect was negative in immediate testing, but positive when the test was given after a delay filled with the study of another, unrelated list.

We have also shown that the SAM model correctly predicts the main results of these experiments. The prediction of the positive effect with randomly chosen cues, in particular, was surprising since we had originally assumed that the model would predict a negative effect. This clearly shows that in the case of a complicated phenomenon such as the part-list cuing effect, one should not rely on intuitive speculation or (as we did) on simplified assumptions. Rather, the lesson that we learned is that one should always run simulations of the actual model to verify one's intuitions. This may sound trivial, but the history of the theoretical analysis of the part-list cuing effect shows that it is often disregarded. Thus, in his review of the literature, Nickerson (1984) incorrectly ascribes a number of assumptions to the SAM model, assumptions that are in fact not made, as would have been clear had the actual model been examined in more detail. For example, Nickerson (1984, p. 549-550) argues that the SAM model makes the assumption that both groups sample the same number of clusters and that the only justification for this assumption is the fact that it is needed to explain the effect. However, the SAM model does not *assume* a priori that both groups sample the same number of clusters. Rather, this may, in practice, often be the result of a much more basic assumption in SAM, namely that there are "costs" involved in retrieving items, and that retrieving items that were already retrieved in prior retrieval cycles increases the likelihood that subjects will stop the retrieval process for these items or those in the same category.

Finally, we have also shown (once again) that, contrary to what is often said, the SAM model does not have so many degrees freedom that it may predict any result. As Raaijmakers and Shiffrin (1981a) already demonstrated, the prediction of a negative part-list cuing effect in regular free recall holds for almost all reasonable values of the parameters and is a consequence of the basic structure of the model. In this study, we also found that the prediction of a positive cuing effect for random cuing of lists of paired-associates is obtained for almost all values of the parameters. Moreover, in those cases, where the effect does depend on the value of a parameter (as was the case for the contextual strength parameter), it is possible to set up experiments to test such a dependency (as we did with the delay manipulation).

All in all, the present results demonstrate that the SAM model is still the only viable explanation for the part-list cuing effect and that results that were thought to be inconsistent with the model, can in fact be shown to follow from its assumptions. It is also clear, however, that, even though the part-list cuing effect itself may not be a puzzle anymore, analyzing why the SAM model predicts this effect remains a puzzling affair due to the fact the phenomenon is dependent on many details of the experimental procedure, making it impossible to use one's intuition to generate predictions. Rather, one should always run the actual simulations. This also shows why quantitative modeling is essential and should not just be considered a personal hobby of a some mathematical enthusiasts.

Summary

The SAM model proposed by Raaijmakers and Shiffrin (1981a) provides an explanation for the part-list cuing effect first observed by Slamecka (1968). In this chapter we present experimental evidence that supports some of the crucial predictions of the SAM model, namely that the part-list cuing effect depends on the relation of the cue items to the stored associative structure and the ability of subjects in the noncued condition to self-generate cues. In Experiment 1 an unexpected positive effect was obtained when the study list consisted of paired-associates rather than single words. This result was replicated in Experiment 3 for a list composed of small-sized categories. This experiment also showed a negative part-list cuing effect when two items from each of five categories were given as cues, whereas a large positive effect was obtained when one cue from each of ten categories was given. In addition, Experiment 2 showed a reversal of the cuing effect depending on whether the recall test was given immediately or after a delay with study of an unrelated list.

Simulation results showed that the SAM model could handle this pattern of results quite well. Other explanations (Roediger & Neely, 1982) that have been given for the positive cuing effects that are sometimes obtained (and were replicated in the present experiments) do not seem to be able to explain the change of the effect from negative to positive, depending on the overall level of recall. Thus, the present results provide strong support for the sampling-bias hypothesis that has been proposed by Raaijmakers and Shiffrin (1981a) as an explanation for the part-list cuing phenomenon.

References

Basden, D.R. (1973). Cued and uncued free recall of unrelated words following interpolated learning. *Journal of Experimental Psychology*, *98*, 429-431.

Basden, D.R., Basden, B.H., & Galloway, B.C. (1977). Inhibition with part-list cuing: Some tests of the item strength hypothesis. *Journal of Experimental Psychology: Human Learning and Memory*, *3*, 100-108.

Blake, M., & Okada, R. (1973). Intralist cuing following retroactive inhibition of well-learned items. *Journal of Experimental Psychology*, *101*, 386-388.

Hudson, P.T.W. (1982). *Preliminary category norms for verbal items in 51 categories in Dutch*. Technical Report 82FU05, Department of Psychology, University of Nijmegen, Nijmegen, the Netherlands.

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

Nickerson, R.S. (1984). Retrieval inhibition from part-set cuing: A persisting enigma in memory research. *Memory & Cognition*, 1984, *12*, 531-552.

Park, N.W. (1980). *Superadditivity of retrieval cues as a function of encoding operations*. Unpublished doctoral dissertation, University of Toronto.

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

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. (1981a). Search of associative memory. *Psychological Review*, *88*, 93-134.

Raaijmakers, J.G.W., & Shiffrin, R.M. (1981b). Order effects in recall. In J.B. Long & A.D. Baddeley (Eds.), *Attention and Performance IX.* (p. 403-415). Hillsdale, NJ: Lawrence Erlbaum Associates.

Roediger, H.L. (1993). Learning and memory: Progress and challenge. In Meyer, D.E. & Kornblum, S. (Eds.), *Attention and Performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience.* (pp. 509-528). Cambridge, MA: The MIT Press.

Roediger, H.L., & Neely, J.H. (1982). Retrieval blocks in episodic and semantic memory. *Canadian Journal of Psychology*, *36*, 231-242.

Slamecka, N.J. (1968). An examination of trace storage in free recall. *Journal of Experimental Psychology*, *76*, 504-513.