Journal of Experimental Psychology
1969, Vol. 80, No. 1, 95–100

INTERFERENCE IN SHORT-TERM MEMORY[1]

GERALD M. REICHER[2]
University of Oregon

ELIZABETH J. LIGON
University of Michigan

AND CAROL H. CONRAD
University of Oregon

Two experiments investigated the locus of the effects of acoustic similarity in short-term memory. In the first, acoustic similarity of words making up the to-be-remembered list was varied independently of the acoustic similarity between the correct and incorrect response alternatives. The results indicate that *S*s are more likely to choose the alternative which is similar to the greatest number of items in the list. A second experiment using a one-alternative yes/no procedure attempted to insure that the effects were not due to a guessing strategy and to allow the use of the signal detectability measure of sensitivity. Similarity within the list was shown to increase both hit and false-alarm rates resulting in an overall loss of sensitivity.

It has become clear that certain types of interference are effective in producing decrements in short-term memory. Conrad (1964) has shown that acoustic similarity of visually presented stimulus material causes decrements in performance in immediate memory tasks. Wicklegren (1965) has shown that the acoustic similarity of the material interpolated between stimulus material and the cue for recall affects short-term forgetting. Keppel and Underwood (1962) have shown that information presented prior to an item which is to be recalled decreases performance in a short-term memory task. A pure trace decay theory of forgetting needs additional assumptions to handle these effects. The assumptions generally made are that similar items create more noise in the system from which selected items must be retrieved than do dissimilar items. Within this framework, two possibilities have been suggested (Posner & Konick, 1966). A response competition (or trace comparison) theory might assume that the interference effects of acoustic similarity do not affect the rate that information decays, but only the difficulty of making the discrimination between the correct item to be remembered and the incorrect response alternatives. Thus, for any given trace strength, a B would be more difficult to distinguish from a D than from a Q. The list interference or "acid bath" model (Posner & Konick, 1966) suggests that the rate of trace decay is faster as a result of increased similarity (analogous to using stronger acid in the process of trace decay) or of increased stimulus size (analogous to using more acid). Posner and Konick's data seem to indicate that interference is active during the retention interval rather than at the time of recall and depends more on time in store than on the strength of the trace. These results support their acid bath model in favor of a response competition model. Their conclusions, however, rely primarily on the lack of a significant interaction, and although there is support from other data in their experiments, it would seem that a more direct test of the relative effects of list interference and response competition would be desirable.

The two important features of the acid bath model (similarity and decay effects) are much like those of a model proposed to explain generalization (Shepard, 1958) and forgetting in a recognition memory task (Shepard, 1961; Shepard & Teghtsoonian, 1961). This model suggests that the

[1] This research was supported by the Advanced Research Projects Agency of the Department of Defense and was monitored by the Air Force Office of Scientific Research under Contract No. F44620-67-C-0099. The authors would like to thank Michael Posner and Richard Haller for their suggestions.

[2] Requests for reprints should be sent to Gerald M. Reicher, Department of Psychology, University of Oregon, Eugene, Oregon 97403.

generalization gradient is a function of spatial proximity of central nervous system representations of stimulus or response events. In some given period of time (or alternatively in some number of trials or presentations), there is some probability that the trace representation of a particular stimulus will migrate to the location of the representation of another stimulus. This probability varies directly as a function of the distance between the two representations. The decay feature of the model is represented by one minus the probability that a trace representation will remain active in memory during a given time. Thus, the strength of a particular trace is a function of the likelihood that a stimulus has activated its internal representation, the likelihood that a trace has migrated to its internal representation, and the likelihood that the traces from these two sources have persisted. The difference between the acid bath and the generalization models is primarily that the acid bath model does not specifically allow for incrementing trace strength of items which have not been presented. The generalization model explains the effects of similarity by these increments of trace strength of incorrect items during and following introduction of similar material. The acid bath model explains similarity effects by decrements in strength of the correct item following introduction of similar material. Both models would predict that list interference would be an important factor in forgetting.

Finally, response competition and list interference are much like the interference theory constructs of response competition and unlearning. The authors will not attempt to choose between interference theory and the decay theories presented above in this paper, but instead will concentrate on the distinction between response competition and list interference.

One way of accomplishing a direct test of the notions of response competition and list interference would be to provide *S*s with stimulus lists which vary in the amount of acoustic similarity (list interference) while independently providing recognition response alternatives which vary in acoustic similarity. Thus, list similarity would presumably not be confounded with response competition because *S* need consider only those response alternatives which are provided. The list interference versions of decay theory would predict that the high-similarity lists would be more interfering so that poorer performance should be expected in the recognition task in this condition. The response competition version would suggest that similarity between the response alternatives would be the most important factor in producing forgetting. Notice that the generalization model can explain some amount of response competition as operationally manipulated in this experiment. Competition at the time of responding should occur as a function of introduction of response alternatives which might be mistaken for items which had appeared in the list. Of course, it is possible that both types of interference will be important in short-term memory.

Experiment I

Method

Subjects.—The *S*s were 24 students at the University of Oregon obtained from the employment office. They were paid $1.50 for their participation.

Procedure.—Each *S* was given the following directions before beginning the experiment:

> You will be listening to a series of words on the tape recorder, using these earphones. You are to repeat aloud each word as soon as you hear it. It is important that you repeat as many of the words as possible. Remember, you are to say the word as soon as you hear it.

Each *S* was then given four-eight practice trials to become accustomed to the task of repeating (shadowing) words. Practice trials were terminated as soon as *S* had correctly shadowed one list of 16 words or had shadowed two lists with only two errors. The rate of presentation of the words in the practice and experimental sessions was 100 words per minute. Then each *S* shadowed 72 lists of words, varying in length from 16 to 20 words per list. At the end of each list *E* stopped the tape recorder and asked *S* to turn over a card with two words printed on it and indicate, by pointing, which of the two words was in the list that he had just repeated. Each card had one correct and one incorrect response. Responses were recorded by *E* and *S* was unable to see his previous responses. Shadowing errors were also recorded by *E*.

Materials.—Six different sets of 36 lists were used, with eight *S*s shadowing Sets 1 and 2, eight shadowing Sets 3 and 4, and eight shadowing Sets 5 and 6. Within each group of *S*s, half began with Sets 1, 3, or 5 and half began with Sets 2, 4, or 6. The *S*s were given a short rest period between the two sets of lists.

Within each set, 12 lists contained 16 words, 12 contained 18 words, and 12 contained 20 words. The critical word, i.e., the word which *S* was required to identify as being in the list, was always thirteenth from the end of the list, placing it in the fourth, sixth, or eighth position depending on the length of the list.

All of the words had only one syllable and all were words commonly used in the English language. Each list had two groups of five or six words which rhymed with or began with the same letter as a key word. For example, the words *neat, heat, seat, nod,* and *nut* constitute a group of five, with *neat* as the key word; *mad, sad, pad, back, bell,* and *book* constitute a group of six with *bad* (not in the list) as the key word. The words of a group were randomly placed in the last 13 positions in a list. The remaining words in a list did not rhyme with or begin with the same sound as any other word in that list.

Lists were divided into six different types, based on the response alternatives from which *S* had to choose the word in the list. Thus the correct alternative was acoustically similar to (confusable—rhymed—with) a group in the list (Cr) or not (Cn), the incorrect alternative was acoustically similar to (confusable with) a group in the list (Ir) or not (In), and the two alternatives were similar to (confusable with) one another (Br) or not (Bn). The six types of lists were CrIrBn, CnInBn, CrIrBr, CnInBr, CrInBn, and CnIrBn.

Set 1 contained 36 lists with six of each of the list types in random order. In Sets 3 and 5 the lists and to-be-remembered words were changed slightly so that they would represent different conditions than they did in Set 1. Sets 2, 4, and 6 contained the same lists as Sets 1, 3, and 5, respectively, except that the incorrect response alternatives in the odd numbered sets became the correct response alternatives in the even numbered sets. Each *S* received two of these sets and thus contributed up to 12 observations on each type of list. Data are not presented from trials on which the critical item was not shadowed correctly or from trials on which *S* intruded the incorrect response alternative during shadowing.

Results and Discussion

Table 1 shows the overall proportion of errors for those trials on which *S* shadowed the critical item correctly. Significance of the results was assessed by a single factor analysis of variance with repeated measures. An arcsin transformation was performed on the raw data—proportions of errors—for the purposes of statistical testing. After determining the main effect, $F\ (5, 115) = 31.00$, $p < .001$, the Newman-Keuls method was used for testing the differences between the means of the transformed data.

TABLE 1

PROPORTIONS OF ERRORS AND STATISTICAL CONTRASTS AS DETERMINED BY THE NEWMAN-KEULS METHOD OF MULTIPLE COMPARISONS

CnIrBn	CrIrBn	CrIrBr	CnInBr	CnInBn	CrInBn
.615	.345	.293	.250	.166	.107
	———	———	———	- - -	- - -
			- - -	- - -	
		- - -	- - -	- - -	

Note.—For those conditions which are not underlined by a common line, $p < .005$. For those conditions which lie over the end points of a common broken line, $p < .05$.

Since each *S* had two replications of the design, the data were broken down by first half and second half of the experiment in an attempt to evaluate progressive effects. The ordering among the groups was the same for the first and second halves of the experiment with one exception. Performance on Cond. CnInBr was worse in the first half and better in the second half of the experiment than performance on Cond. CrIrBr. Overall, performance on the first half was somewhat better than performance on the second half.

Examination of Table 1 shows that the data do not support the most obvious predictions of the acid bath model taken by itself. For example, memory for the critical letter in Cond. CnIrBn is not easier than in Cond. CrIrBn, but rather the opposite seems true. In fact, performance on Cond. CnIrBn is poorer than chance performance ($p < .05$ assessed by a sign test [1] comparing each *S*'s proportion of correct responses with .50). The same tendency appears in the comparison between CnInBn and CrInBn. Thus, having an item to be remembered embedded in similar items in a recognition task does not reduce the chance that the item will be recognized. Finally, performance on the condition which should have a great deal of interference according to the acid bath

[1] All sign tests reported in this paper are two-tailed.

model, CrIrBr, is not significantly worse than the condition which should have little or no interference, CnInBr. The trace comparison model fares little better. The nonsignificant difference between CrIrBr and CrIrBn is in the wrong direction. It was found, however, that the difference between CnInBr and CnInBn is in the right direction.

The data do seem to support the notion that *S*s will choose the alternative which is most similar to the items in the list as would be predicted by Shepard's (1961) generalization model. Thus, the easiest list is one in which a group of filler items are all similar to the critical item (CrInBn); the most difficult is where a group of filler items are not similar to the critical item but rather similar to the incorrect response alternative (CnIrBn). If one makes the further assumption that the amount of tendency to respond to a particular alternative on the basis of identity or similarity reaches some limit such that additional similarity provides less and less tendency to respond, most of the rest of the data of this experiment can be accounted for. The intended implication of this last statement is that similar items will add more tendency to respond when they are not similar to the critical item than when they are. The relevant contrasts are that Group CrIrBn does less well than Group CnInBn. Also, Group CnInBr does slightly (not significantly) better than Group CrIrBr.

Thus, the impression which one might get from these data is not that information is fading away until performance is at a chance level. Rather, for this forced-choice recognition task, performance seems to be a function of the amount of information in the list which might confuse with the correct or incorrect response alternatives. This notion of building up of confusion rather than decaying of information is emphasized by the fact that performance is poorest on those conditions in which the incorrect alternative is similar to items in the list; in one such condition, performance is worse than chance.

There are, of course, alternative explanations. It is possible, e.g., that *S*s chose the way they did because they are forced to make one choice or the other. Since there was no way that *S* could say that he had not seen an item, having forgotten the critical item, he might have chosen the alternative which rhymed with the greatest number of words in the list. To test this hypothesis, a second experiment was run.

Before consideration of Exp. II, one might question whether a memory experiment purporting to be about acoustic similarity effects would not be hopelessly confounded by perceptual errors inherent in auditory presentation of the material. A perceptual explanation of these data might be that the acoustic similarity of the material caused *S*s to misperceive the to-be-remembered word and thus shadow it incorrectly; *E* might then misperceive *S*'s statement of the word, in accordance with his expectancy to hear the correct word, and fail to eliminate the trial. Table 2 shows the breakdown of the shadowing errors relevant to perceptual explanations of these data. The inferences made on the basis of the data presented in this table assume that *E* caught at least a reasonable proportion of the shadowing errors. The most important column with respect to the above perceptual hypothesis is the one relating shadowing errors on to-be-remembered words to conditions. It can be seen that these error rates and the differences among the groups are too small to account for the memory data.

The types of errors in this experiment also argue strongly against the perceptual hypothesis presented above. According to the

TABLE 2

SHADOWING ERRORS

Condition	Proportion of critical items shadowed incorrectly	Proportion of trials on which the incorrect alternative was intruded
Exp. I		
CnIrBn	.045	.007
CrIrBn	.063	.003
CrIrBr	.066	.010
CnInBr	.056	.000
CnInBn	.066	.000
CrInBn	.045	.000
Exp. II		
Cr	.060	—
Cn	.042	—
Ir	—	.000
In	—	.005

perceptual hypothesis, one would expect that the Cr conditions would show the greatest error rates because *S* would fail to correctly perceive and thus fail to remember the items in the high-similarity lists. However, just the opposite error tendency was observed. The *S*s tended to do badly on the Ir conditions in which they falsely recognized an item which had not been presented but which was similar to items in the list. Thus the perceptual hypothesis would be forced to say that these data are the result of *E*'s misperception of *S*'s shadowing intrusions of the word which was to be the incorrect alternative. Table 2 shows these error rates, and again the rates are much too low to account for the memory data.

EXPERIMENT II

This experiment was designed to see whether the same results would be obtained if *S*s were allowed to say that they did not see a particular response alternative. The only difference between Exp. I and II is that a single yes/no response alternative was presented to *S* in Exp. II. Thus, if the critical item was no longer in store at the time of the test for retention, *S* could say "No," the response alternative was not in the list just presented.

Method

Twelve *S*s were drawn from the same source as in Exp. I. The only other difference between the two experiments was the response measure. For a given list, *S* was given one of the two alternatives used for that list in the forced-choice procedure of Exp. I. Another *S* got the other of the two responses. This division was done in a way which insured that each *S* had the same number of similar and dissimilar response alternatives. The *S* responded "yes" or "no" to each list and gave a confidence judgment 1 if sure, 3 if guessing, and 2 if somewhere between.

This change to a yes/no procedure reduces the number of conditions to four (Cn, Cr, In, Ir) since only one word is given as a response probe and there is no chance for similarity between items.

Results and Discussion

The differences among all groups were significant by a sign test of within-*S* comparisons on the four conditions ($p < .01$). The ordering of the groups was the same for the two halves of the experiment; overall performance was slightly better on the first half. The probability that an *S* will respond "yes" to the response alternative is increased as a function of identity and of similarity to list members. Thus, Cond. Cr shows more correct recognitions than Cond. Cn (68% vs. 52%). The increase in hit rate as a function of similarity is compensated for by an increase in false recognitions—saying "yes" when the alternative was not in the list—in the Ir (40%) condition as compared with the In (7%) condition. This difference in false-alarm rates is observed even when *S*s give the highest confidence rating indicating that they are sure of their response (Ir 17.8%, In 2.3%, $p < .05$ assessed by a sign test). Thus the increase in false alarms or confusions does not seem to be the result of a guessing strategy.

Since these data allow us to compute independent hit and false-alarm rates, d', the measure of sensitivity in signal detectability analysis, was obtained (Swets, 1964, Appendix I). The relative frequency of saying "yes" for the Cr and Cn lists was taken as the hit rate for the high- and low-similarity lists, respectively. The relative frequency of saying "yes" for the Ir and In lists was taken as the false-alarm rate. The d' was higher for low-similarity lists than for high-similarity lists for every *S*. The means of the individual *S*s' d' values were 1.80 and .82, respectively. The tabled values for d' given hit and false-alarm rates assume that the signal and noise distributions are normal with equal variances. In an attempt to technically justify the use of this procedure, ROC plots were made on normal–normal paper using the confidence data to estimate different criteria. This procedure should be carried out separately for each *S*, but there was not enough data to do this so that the data for all *S*s were grouped for purposes of this analysis. An informally fit straight line provides a fairly good fit for the data for both the high- and low-similarity conditions, indicating that to the extent that the estimations of successive criteria are accurate, the underlying distributions of signal plus noise and noise alone are close to normal. The lines are not parallel to each other or to the diagonal so that the variances of the distributions of noise and of signal plus noise cannot be considered equal. This suggests that the estimates of *d'* made above are biased by the ratio of the variances. According to this analysis, the variance of the distribution of signal plus noise was larger than the variance of the distribution of noise alone; this difference was greater in the low-similarity condition. If one takes the intersect value of *d'*, the ratio of the variances does not enter the calculation so that these estimates of *d'* should be free from that source of bias. The values are .64 for the high-similarity condition and 1.01 for the low-similarity condition. Thus, in the sense implied by this measure of sensitivity, it would seem as if list similarity causes a loss of information, as would be suggested by the acid bath and the generalization models described above.

The fact that performance on the high-similarity condition showed a higher hit rate and a lower *d'* than performance on the low-similarity condition indicates that *S*s used a more stringent decision criterion for the low-similarity material than for the high-similarity material.

Conclusions

The results of this experiment indicate that list interference is a factor in short-term memory. The authors have emphasized the building up of confusion rather than the weakening of the to-be-remembered item as an explanatory concept. While the notion of incrementing of classes of responses, including incorrect responses, does gain support from these data, this in no way eliminates the possibility that weakening of the trace as a function of time or similarity also occurs.

References

Conrad, R. Acoustic confusion in immediate memory. *British Journal of Psychology,* 1964, **55,** 78–84.

Keppel, G., & Underwood, B. J. Proactive inhibition in short-term retention of individual items. *Journal of Verbal Learning and Verbal Behavior,* 1962, **1,** 310–329.

Posner, M. I., & Konick, A. E. On the role of interference in short-term retention. *Journal of Experimental Psychology,* 1966, **72,** 221–231.

Shepard, R. N. Stimulus and response generalization: Deduction of the generalization gradient from a trace model. *Psychological Review,* 1958, **65,** 242–256.

Shepard, R. N. Application of a trace model to the retention of information in a recognition task. *Psychometrika,* 1961, **26,** 185–202.

Shepard, R. N., & Teghtsoonian, M. Retention of information under conditions approaching a steady state. *Journal of Experimental Psychology,* 1961, **62,** 202–222.

Swets, J. A. (Ed.) *Signal detection and recognition by human observers.* New York: Wiley, 1964.

Wicklegren, W. Acoustic similarity and retroactive interference in short-term memory. *Journal of Verbal Learning and Verbal Behavior,* 1965, **4,** 53–61.

(Received May 27, 1968)