Journal of Experimental Psychology
1973, Vol. 100, No. 1, 73–82

RECOGNITION EFFECTS OF STUDY ORGANIZATION AND TEST CONTEXT

LARRY L. JACOBY[1] AND REGINALD L. HENDRICKS

Iowa State University

Effects of study list spacing of category instances and recognition test order were investigated in 2 experiments. Frequency of study and new test items as category name associates were also included as variables in Experiment II. In both experiments, grouping items by category on the recognition test enhanced performance if instances had also been grouped during study. In Experiment II, low-frequency study items were recognized better than high-frequency ones. The interfering effect of increasing the frequency level of new test items was maximal when *old* items were high frequency and not grouped by category on the recognition test. Results were discussed with reference to the role of organization and context in recognition memory.

The problem in recall is often not one of storing information but rather one of finding the information in memory once it has been stored. There is now an abundance of research that demonstrates that recall failures are often due to failures in retrieval (e.g., Tulving & Pearlstone, 1966). Is a similar failure in retrieval a possible cause of recognition errors? McCormack (1972) recently reviewed the literature and concluded that there is no evidence of complex search processes operating in recognition memory. Recognition performance was viewed as a relatively pure measure of storage; the number of cues provided by a recognition test was considered to be such that search and retrieval are unnecessary or trivial. Several other investigators have drawn similar conclusions (e.g., Kintsch, 1970; Murdock, 1968). The distinction between storage and retrieval effects has been used quite often in attempts to specify the locus of organization effects. One possibility is that organization serves to aid retrieval without influencing the storage of individual items (Slamecka, 1972). The implication is that recognition performance should be uninfluenced by organization. A second possibility is that organization influences the storage of items without the influence being such that recognition is affected (Kintsch, 1970). Two major questions are raised: (*a*) Does organization influence recognition performance? (*b*) Does recognition involve search and retrieval? If so, what is being retrieved?

Organization is typically used as a theoretical construct that can be measured only indirectly; clustering of related items in free recall is one of the organization measures used most often (Shuell, 1969). If category clustering exceeds a chance level, it is taken as evidence that items were organized by category. The level of clustering is also assumed to reflect the influence of variables on organization. For example, the finding that blocking category instances during study increases clustering can be taken as evidence of an effect of study order on organization. A high correlation between clustering and recall is usually found, so that it is tempting to conclude that both are a result of relationships formed among items in storage. That is, it might be concluded that organization has had the effect of unitizing sets of items in storage (Mandler, 1967; Tulving, 1968a). However, clustering does not necessarily rule out the possibility of items being stored independently (Slamecka, 1972). The conclusions that can be drawn from the relationship of clustering and recall are limited by its correlational nature.

Investigations of recognition memory have provided firm evidence of an effect of organization in storage. One source of evidence comes from studies employing

[1] Requests for reprints should be sent to Larry L. Jacoby, Department of Psychology, Iowa State University, Old Botany Hall, Ames, Iowa 50010.

homographs (e.g., Light & Carter-Sobell, 1970). Recognition performance is higher if study and test contexts are similar, so that the meaning given to a homograph is the same on the 2 occasions. It seems necessary to conclude that the context was stored with the item during study or at least served to determine the meaning that was encoded. A similar manipulation has been performed with the recognition test of a categorized list (Jacoby, 1972). Recognition was found to be higher when items were grouped by category on the recognition test, and *old* items were tested in the same order as they had been studied. This result complements the free-recall finding of clustering by category with items in a cluster being recalled in approximately the same order as they were studied (Kintsch, 1970). The recognition effect of test order leads to the conclusion that relationships among category instances were stored and input order information was preserved. It seems impossible to account for these results while maintaining the position that category instances were stored independently.

It appears that organization does influence recognition, and the manipulations of the recognition test are a fruitful means of investigating this influence. It might also seem reasonable to conclude that search and retrieval are involved in recognition. The object of retrieval might be the larger encoded unit in which a test item is represented (Horowitz & Manelis, 1972; Tulving, 1968b). Retrieval requirements would then depend on the similarity of study and test encodings of an item. When encodings are identical, search would be essentially eliminated. When an item has been encoded as a portion of a larger unit during study, recognition of the item presented alone or in a different context during the test might require retrieval of the larger unit. Prior experiments have found that recognition is higher when instances of a category are blocked during study rather than distributed randomly throughout the list (D'Agostino, 1969; Jacoby, 1972).

The present experiments were designed to more carefully investigate the effects of spacing the presentation of category members during study. Levels of spacing in the first experiment are comparable to those employed by Glanzer (1969, Experiment I) in his study of free recall. Glanzer found that recall probability declined with increases in spacing of related items. Manipulations of recognition test order were also employed in the present experiments. Grouping items by category on the recognition test should enhance performance if items were organized by category during study. A second purpose of Experiment I was to assess relationships based on serial order. When category instances are widely spaced, adjacency of study position might serve as an alternative basis of organization.

Experiment I

Method

Materials and procedure. The 4 most frequently reported instances were selected from each of 36 categories listed in the Battig and Montague (1969) norms. Words that held an odd-numbered frequency rank in the norms were employed as study items, while those holding even-numbered ranks served as new distractor items for the recognition test. Thus, a study list contained 72 words, 2 instances each of 36 different categories. Instances of a given category were either presented in contiguous study serial positions or separated by the presentation of 1, 3, or 11 intervening items. The spacing of category instances did not vary within a list.

Three test lists differed only with regard to the order in which items were tested. To construct a random test, *old* and *new* items were randomly assigned to test positions with the restriction that 2 items from the same category could not occur consecutively. Construction of the categorized test was identical to that of the categorized-ordered test employed by Jacoby (1972). Both *old* and *new* items from the same category were grouped so that they occupied consecutive positions in the test list. Within a category, *old* items were tested in the same relative order as they had been studied, although *new* words might intervene between their test occurrences. Items were randomly assigned to test positions with these restrictions. The order in which categories were tested did not correspond to that in which they were studied; category test order was randomly determined. To construct an adjacent test list, consecutive words in the study list were grouped so as to form dyads; words in Positions 2 and 3 formed a dyad as did those in Positions 4 and 5, 6 and 7, etc. Although they were not adjacent during study, words in Positions 1 and 72 were included in a dyad. (Had dyads been formed from Positions 1 and 2, 3 and 4, etc., the adjacent and categorized test lists would have been identical when study spacing was 0.) Distractor items were chosen so that if the *old* words in a dyad were the first and third most frequent items in their respective categories, the *new* items were the second and fourth most frequent. Dyads, along with the appropriate distractors, were grouped into sets with the test order of these sets being random. Within a set, the relative order of *old* test words was the same as during study. A separate adjacent test list was constructed for each level of study spacing.

Test lists were prepared as mimeographed booklets. The first page of the booklet was blank with the exception of a sentence that informed *S*s that they were not to turn that page until instructed to do so; the second page contained instructions for the recognition test. Instructions stated that each of the following pages would contain a list of words, a portion of which had occurred in the study list. The *S*s were instructed to respond to each item in the order that it was listed, circling *old* or *new* depending on whether or not the word had occurred in the study list and then rating their confidence in that judgment on a 5-point scale. The scale was to be interpreted as extending from extreme confidence to complete uncertainty. Each of the last 6 pages of the test booklet contained a single column of 24 words, one half of which had occurred in the study list.

Study items were presented visually at a rate of 2 sec/item. After presentation of the last item, *S*s were instructed to turn the first page of the test booklet. Prior to reading the test booklet instructions, *S*s were not informed about the nature of the retention test. There was no time limit placed on the recognition test.

Design and subjects. Four levels of spacing (0, 1, 3, and 11) were factorially combined with 3 levels of test (random, categorized, and adjacent) to form 12 between-*S*s experimental conditions. The *S*s were 168 students who were enrolled in an introductory psychology class and participated for course credit; 14 *S*s were assigned to each of the 12 experimental conditions. The *S*s were tested in small groups that ranged in size from 3–5 people. All *S*s tested in a given session received the same level of spacing; each level of test was represented by at least 1 *S* in each session. With these restrictions, the assignment of *S*s to conditions was random.

Analyses. Two recognition measures were employed. A difference measure was obtained for each *S* by subtracting the number of *old* responses given to *new* distractor items (false alarms) from the number of *old* responses given to study items (hits). The signal detection model was employed to obtain d' as a second measure of recognition. In general, *S*s only used the extreme points on the confidence scale, and analyses of these data added no new information. For these reasons, analyses involving confidence judgments are not included in the results.

A third measure reflects the joint probability of recognizing both *old* items from a given category. For each *S*, the number of categories from which both *old* items were recognized was multiplied by 2 and divided by the total number of hits. If *S* recognized both items from 18 categories and had a total of 48 hits, his score would be .75. This score is the proportion of items in each category that were recognized, given recognition of at least 1 item in the category. The rationale behind this measure is identical to that underlying the use of item per category scores in studies of recall (e.g., Tulving & Pearlstone, 1966). A higher proportion of items recognized within a category should result from increasing organization. Within limits, the proportion score can vary independently of the probability of a hit. Constraints on this independence are the same as those on item per category scores in recall.

Results and Discussion

Recognition measures for each combination of conditions are summarized in Table 1. The measures presented are: probability of a hit; probability of a false alarm (FA); difference between probabilities of

TABLE 1

RECOGNITION MEASURES FROM EXPERIMENT I

Spacing	Test order														
	Random					Adjacent					Categorized				
	P(hit)	*P*(FA)	*D*	d'	*PPC*	*P*(hit)	*P*(FA)	*D*	d'	*PPC*	*P*(hit)	*P*(FA)	*D*	d'	*PPC*
0	.80	.07	.73	2.49	.85	.76	.06	.70	2.41	.83	.88	.07	.81	3.04	.88
1	.78	.12	.66	2.23	.84	.80	.11	.69	2.22	.82	.87	.07	.80	2.97	.88
3	.68	.12	.56	1.92	.72	.76	.14	.62	2.19	.77	.71	.14	.57	1.66	.74
11	.70	.20	.50	1.49	.72	.81	.17	.64	1.95	.80	.76	.14	.62	1.97	.78

Note. Abbreviations: *P*(hit) = probability of a hit; *P*(FA) = probability of a false alarm; *D* = difference between probabilities of a hit and a false alarm; and *PPC* = proportion of items per category recognized, given recognition of at least 1 instance of the category.

a hit and false alarm (*D*); d' means; and proportion of items per category recognized, given recognition of at least 1 instance of the category (*PPC*).

Spacings of 0 and 1 produced nearly equivalent performance, as did spacings of 3 and 11. The probability of a hit decreased with spacing, while the probability of a false alarm increased. In general, the categorized test produced a higher probability of a hit and lower probability of a false alarm than did the random test; the advantage of the categorized test did not hold when category instances were separated by 3 intervening items. Performance in the adjacent test condition was intermediate to that of the other 2.

Analyses of difference and d' scores provided general support for the description of results given above. The same effects were found to be significant in both analyses, so only the results of the difference analysis will be reported. The effects of spacing, F (3, 156) = 13.94, $p < .001$, and test order, F (2, 156) = 4.66, $p < .05$, were both significant. Newman-Keuls tests were employed for individual comparisons. These comparisons revealed that spacings of 0 and 1 produced better recognition than did spacings of 3 and 11 (all ps $< .01$); other differences among spacing conditions were nonsignificant. Performance was higher when the test was categorized rather than random ($p < .01$). Recognition in the adjacent test condition did not differ from that in either of the other test conditions. The interaction of spacing and test condition was not significant.

Organization was expected to decline with increases in spacing so that there would be a lower probability of recognizing both instances of a category. The probability of recognizing both instances given recognition of one is reflected by the proportion of items per category measure. The proportion of items per category that were recognized did decline with increases in spacing, F (3, 156) = 12.55, $p < .001$, and the pattern of this decline was identical to that observed in overall recognition performance; spacings of 0 and 1 produced nearly identical performance, as did spacings of 3 and 11. The correspondence of overall recognition and the items per category measure did not hold when the effects of test order were examined. Categorizing the recognition test enhanced overall performance but did not influence the proportion of items per category that were recognized. This result is similar to the effect observed when cued and free recall of a categorized list are compared (Tulving & Pearlstone, 1966). Presenting category names as cues increases the total number of words recalled but does not influence the number of items per category recalled. As in recall, it seems that the recognition test manipulation serves to provide access to additional higher order units that were established during study.

What has been described as organization may actually be no more than the effects of rehearsal; spacing of category instances might simply influence the directing of rehearsal. When category instances are presented near to one another during study, the first instance of a category might be rehearsed during presentation of the second, and this rehearsal might be more effective than that produced when instances are widely spaced. The implications of this interpretation are: (*a*) The first instance of a category should be recognized more often than the second, and (*b*) the recognition advantage of the first instance in a category should decline with increases in spacing. An analysis of hits did reveal that the first instance of a category was recognized with a higher probability (.80 vs. .76) than was the second, F (1, 156) = 16.09, $p < .001$. There was also a tendency toward a Presentation Position $\times$ Spacing interaction, F (3, 156) = 2.54, $p < .10$. However, the interaction was not of the form that might be expected. The probability of recognizing either the first or second instance of a category decreased with spacing, but the decrease was more pronounced for the second instance. As a result, the recognition advantage of the first instance of a category increased with spacing, the opposite of what would be expected from the rehearsal position.

Spacing of category instances apparently influenced the storage of relationships among items. One possibility is that the distance between related words influences the aspects and completeness of meaning that is encoded. When instances of a category are nearly simultaneous during study, the items are compared and relationships are noted, including both similarities and contrasts in meaning. The result is much like an association that is stored in addition to the 2 words. Increasing spacing diminishes the amount of information about the first instance of the category that is accessible during the presentation of a second instance of the same category. The effect is much like that of the weakening of an association, but what is meant is that fewer relationships between instances are established. Results of the present experiment suggest that organization did not vary in an all-or-none fashion with spacing. Some relationships were apparently encoded even with the wider degrees of spacing. If they were not, there would be no reason to expect recognition of the first presented instance of a category to be higher than that of the second. The high probabilities of false alarms with spacings of 3 and 11 can be taken as evidence that *S*s encoded category information but less additional information that was of assistance for discriminating between *old* and *new* instances of a category. Even with high degrees of spacing, some relationships were apparently established during study and made more accessible by categorizing the recognition test.

Experiment II

The memory trace can be viewed as a collection of attributes including semantic, acoustic, and orthographic features of a presented word (Bower, 1967; Underwood, 1969; Wickens, 1970). Additional attributes might carry information about factors surrounding the presentation of an item or associations that were formed during study. What determines the selection of attributes for inclusion in the memory trace? It is possible that an item is encoded completely and almost automatically with regard to all attributes (Wickens, 1970). However, it would seem more likely that encoding is influenced by past experience with an item and the context surrounding an item's presentation. Items on a recognition test might be recognized equally well but for different reasons. In some cases, recognition might be on the basis of acoustic or orthographic features, while meaning and relationships established during study allow recognition in other cases.

Category norms have been established by presenting category names and requiring *S*s to provide instances of the categories (e.g., Battig & Montague, 1969). In this way, the frequency of a word as a response to a particular category name is indexed. This frequency might also serve as an index of the likelihood of the category name being included in the encoded version of the category instance. For example, the category name ANIMAL might be very likely to be encoded if the word DOG is presented for study; encoding of the category name would be less likely if AARDVARK was the word that was to be remembered. Categorization might occur with high-frequency items, while acoustic and orthographic information is emphasized in the coding of low-frequency ones. The encoding of acoustic and orthographic information may be relatively context free. That is, recognition based on these types of information may not be influenced by changes in context.

The encoding of meaning is influenced by the context surrounding presentation of an item. An obvious example is the influence of context on the meaning ascribed to homographs (e.g., Light & Carter-Sobell, 1970). A more general effect might apply to the encoding of all semantic information. When related words appear in close succession, relationships might be noted and emphasized in the encoding of the items. Categorization of an item might be more likely if the item occurs with other instances of the same category. For example, AARDVARK might not be encoded as a member of a more general category if it is presented with an unrelated word, but would be encoded if presented with another instance of the animal category. Results of Experiment I suggested that spacing influences the establishment of relationships among items. Category information and other relationships are more likely to be encoded if category instances occur near to one another during study. Regardless of spacing, category membership of high-frequency items is likely to be encoded. Spacing of low-frequency items may influence both categorization and the establishment of additional relationships.

Recognition effects of frequency and context were investigated in the present experiment. Both frequency level and spacing were expected to influence the study encoding of category instances. Effects of test order and frequency level of *new* test items should depend on information stored during study. The test manipulations should have an effect only if category information has been encoded. The effect of categorizing the recognition test should also reflect study organization of category instances; categorizing the test will not aid recognition unless relationships among instances have been established during study.

Method

Design and subjects. Two levels of study spacing (0 and 11) were factorially combined with 2 test orders (categorized and random) to form 4 between-*S*s conditions. Normative frequency of study items (high and low) and *new* distractor items (high and low) were varied within-*S*s. The resulting design was a 2 × 2 × 2 × 2 factorial.

The *S*s were 64 students in psychology classes who volunteered to participate for course credit. Sixteen *S*s were assigned to each of the 4 between-*S*s experimental conditions. The method of assigning *S*s to conditions was identical to that employed in Experiment I.

Materials and procedure. Four words were selected from each of 36 different categories in the Battig and Montague (1969) norms. High-frequency items were the 4 most frequent instances of their respective categories and had a mean total frequency of 320.07 as associates to their category name. Low-frequency instances had a mean total frequency of 13.15. The frequency of occurrence in English (Thorndike & Lorge, 1944) of high-frequency words ranged from 1 in 4,000,000 to AA (median 47.5 per million); that of low-frequency words ranged from no occurrence in the word count to A (median 7.5 per million). The confounding of frequency in the language and associative frequency to a category name is undesirable but appeared to be impossible to avoid when selecting word pools of the size required in the present investigation.

A study list contained 72 words; 2 instances each of 36 different categories. One half of these categories were represented by high-frequency instances, while the other half were represented by low-frequency ones. Nine categories were assigned to each of the 4 test combinations of frequency of *old* and *new*. Within each combination, *old* and *new* items were equal in number and selected from the same categories. Two replications of the basic design were formed by interchanging words assigned the role of study items with those that had served as *new* distractor items.

Study lists for each level of spacing were constructed by the same means as in Experiment I. Construction of test lists was also identical with the exceptions that confidence ratings were not taken, and *S*s indicated their responses by circling *old* items. All other details of the method and procedure were the same as in Experiment I.

Results and Discussion

Recognition measures for each combination of conditions are presented in Table 2; measures were the same as those employed in Experiment I. To simplify presentation, results will be described and discussed separately for the 2 levels of study spacing. This division of results is also justifiable on the basis of expectations described in the introduction of Experiment II. Results of the d' analysis will be reported only when they are at odds with those of the analysis of difference scores. In a last section, results of Experiment II will be compared with those of Experiment I.

Effects with spacing of 0. Organization of items during study was expected when instances of a category were presented contiguously. This organization should be reflected by effects of test order and normative frequency of distractor items. Categorizing the test should not aid recognition unless relationships among instances were established during study. Similarly, the normative frequency of *new* test items should not be an important factor unless category information was encoded during study.

Low-frequency study items were recognized more accurately than were high-

TABLE 2

Recognition Measures from Experiment II

Spacing	Test	High frequency of study items										Low frequency of study items									
		High frequency of *new* items					Low frequency of *new* items					High frequency of *new* items					Low frequency of *new* items				
		P (hit)	*P* (FA)	*D*	*d'*	*PPC*	*P* (hit)	*P* (FA)	*D*	*d'*	*PPC*	*P* (hit)	*P* (FA)	*D*	*d'*	*PPC*	*P* (hit)	*P* (FA)	*D*	*d'*	*PPC*
0	Random	.72	.08	.64	2.69	.84	.75	.03	.72	3.69	.86	.73	.04	.69	3.48	.80	.81	.04	.77	3.88	.80
	Categorized	.80	.05	.75	3.60	.78	.81	.05	.76	3.56	.83	.83	.03	.80	4.27	.80	.88	.02	.86	4.48	.87
11	Random	.75	.09	.66	2.50	.74	.75	.03	.72	3.61	.77	.76	.05	.71	3.61	.74	.80	.08	.72	3.24	.84
	Categorized	.59	.04	.55	3.04	.52	.59	.04	.55	2.94	.47	.71	.02	.69	3.75	.65	.73	.03	.70	3.53	.67

Note. Abbreviations: *P* (hit) = probability of a hit; *P* (FA) = probability of a false alarm; *D* = difference between probabilities of a hit and a false alarm; and *PPC* = proportion of items per category recognized, given recognition of at least 1 instance of the category.

frequency ones, F (1, 60) = 5.92, $p < .05$. However, frequency level apparently did not influence the establishment of relationships among items during study; effects of test manipulations were relatively constant across the levels of study-item frequency. Recognition performance was superior when *new* test items were low frequency, F (1, 60) = 8.08, $p < .01$. Categorizing the recognition test enhanced performance; this effect was most pronounced when high- rather than low-frequency instances served as *new* test items. The main effect of test order approached significance in the analysis of difference scores, F (1, 60) = 3.01, $p < .10$. The d' analysis revealed a significant Test Order × Frequency of *New* Test Items interaction, F (1, 60) = 4.39, $p < .05$. Analysis of the proportion of items per category measure failed to reveal any significant effects.

Categorizing the test did less to aid performance when *new* items were low-frequency instances of a category. This result may have been due to an increase in similarity of *old* items and low-frequency *new* ones; low-frequency instances may not have been encoded as being members of the same category as study items unless the test was categorized. Had it not been for relationships established during study, categorizing the test may have lowered recognition performance.

Effects with spacing of 11. The test manipulations were such that they should influence recognition based on meaning but have little or no effect if recognition were based on acoustic or orthographic features. Acoustic and orthographic features were expected to dominate the study coding of low-frequency items. Thus, manipulating the test of low-frequency study items was expected to have essentially no effect. However, recognition of high-frequency items may be primarily on the basis of meaning, so that manipulations of their test will be effective. Category membership of high-frequency items may be encoded during study, but relatively few additional relationships should be established when category instances are widely spaced. As a result, categorizing the recognition test of high-frequency study items may disrupt recognition performance by increasing the apparent similarity of *new* test items.

Recognition performance was superior when study items were low-frequency instances of a category, F (1, 60) = 12.30, $p < .01$. With low-frequency study items, neither test order nor the frequency level of *new* test items had a substantial effect. Categorizing the recognition test of low-frequency study items reduced the probability of a false alarm, but there was a corresponding decrease in the probability of a hit. With high-frequency study items, both test order and the frequency level of *new* test items were effective variables. Recognition on a random test was superior when *new* items were low frequency; frequency level of *new* test items did not influence recognition on a categorized test. Categorizing the recognition test reduced the probability of correctly recognizing high-frequency study items. The probability of a false alarm was also reduced by categorizing the test when *new* items were high-frequency category instances.

The recognition effect of test order depends on the recognition measure employed. The analysis of difference scores revealed a significant interaction of test order and frequency level of study items, F (1, 60) = 6.38, $p < .025$. Categorizing the test reduced recognition of high-frequency study items but did not influence that of low-frequency ones. In the difference analysis, the effect of test order did not interact with the frequency level of *new* study items. Neither the main effect of test order nor any of its interactions with other variables attained significance in the d' analysis. However, a d' analysis that only included recognition of high-frequency study items did reveal a significant Test Order × Frequency Level of *New* Test Items interaction, F (1, 60) = 8.58, $p < .01$. The condition with low-frequency *new* items enjoyed a recognition advantage only on the random test. Categorizing the test enhanced recognition with high-frequency *new* items but interfered with performance when *new* items were low-frequency category instances.

The analysis of the proportion of items per category measure revealed significant main effects of study-item frequency, F (1, 60) = 5.94, $p < .05$, and test order, F (1, 60) = 46.97, $p < .001$. The proportion of items per category that were recognized was greater when study instances were low frequency. Categorizing the test reduced items per category recognition, and there was a tendency for this effect to be larger when study items were high-frequency instances of a category. However, the interaction of test order and study-item frequency was not significant.

Comparisons of Experiments I and II. Results from the conditions with high-frequency study and *new* test items can be compared across the 2 experiments. When study spacing was 0, performance in the 2 experiments was quite comparable; the magnitude of the test effect was nearly identical across experiments. The major inconsistency in results involves the conditions with a study spacing of 11. With a spacing of 11, the probability of a false alarm was much higher in Experiment I than in Experiment II. The probability of a hit was also higher in Experiment I when the test was categorized, but was lower on a random test. Categorizing the test reduced the proportion of items per category recognized in Experiment II, while leaving it unchanged in Experiment I. The reason for these discrepancies may be differences in number of relationships established between category instances during study. Including low-frequency study items may have obscured the pairwise category structure of the study list so that fewer relationships were established when instances were widely spaced. Results of Experiment I suggest that organization on a category basis is less efficient with a spacing of 11 than with a spacing of 0. Recognition on a random test may not be enhanced and may even suffer with low degrees of organization during study. Categorizing the recognition test emphasizes similarities among items and may disrupt performance unless relationships between category instances have been established during study.

General Discussion

Is the effect of organization or context such that recognition performance is influenced? Prior experiments (e.g., Light & Carter-Sobell, 1970; Tulving & Thomson, 1971) have provided an affirmative answer to this question by finding recognition effects of context change. Results of the present investigation suggest that context also influences the encoding of category instances. Relationships among items are emphasized when instances of a category are presented successively. As a result, the information encoded seems to be substantially different from that encoded when an item is presented with unrelated words. Other investigators have attempted to account for recognition effects of frequency level (Underwood & Freund, 1970) and relationships among items (e.g., Ekstrand, Wallace, & Underwood, 1966) by assuming that implicit associative responses occur when an item is presented. The primary difficulty with these interpretations is that effects of context are largely ignored. Due to context effects, associative strength indexed in a free association situation may be a poor predictor of relationships or other information encoded when an item is presented in a list learning situation.

Is retrieval involved in recognition memory? Mandler (1972) responds to this question by suggesting that retrieval is involved in some situations but not others; retrieval is considered to be involved only when there is a high level of organization of items during study. Results of the present experiments lead to similar conclusions. An item may be recognized on the basis of relationships established with other items during study. These relationships or higher order units are not provided by the presentation of a test item and must be retrieved to aid recognition performance. Categorizing the test can aid recognition by providing access to additional relationships that were established during study. An item may also be recognized on the basis of acoustic, orthographic, or other attributes that are unique to the individual item. The encoding of these attributes may be relatively uninfluenced by context, so that in some situations recognition will be context free. However, recognition on the basis of meaning does appear to be sensitive to changes in context. Categorizing items on the recognition test can disrupt recognition performance if few relationships have been established between category instances during study. One of the oldest laws of recognition memory is that the difficulty of recognition depends on the similarity of study items and *new* distractor items on the recognition test. It appears that the similarity of *new* items is partially determined by the context in which they are presented.

One of the main problems facing a theory of recognition memory is to specify when context effects should be expected. The answer that has been given here is that context effects will occur when items have been organized during study. However, an answer of this type is not a satisfactory one unless the term organization is defined by specifying factors that influence it (cf. Postman, 1972). In this regard, the problem for recognition is the same as that for recall. Some variables that are believed to influence organization have similar effects in recall and recognition. For example, increasing the spacing of related items has been shown to lower recall (Glanzer, 1969, Experiment I) and recognition on a categorized test. Strategies provided by instructions also affect recall and have been shown to influence the recognition effect of context change (Winograd, Karchmer, & Russel, 1971). These similarities in recall and recognition effects of variables are as much in need of explanation as are differences. Encoding processes should be identical prior to a recall and recognition test when *S* is not informed concerning the type of test he will encounter. Free-recall differences in retrieval appear to result from the influence of study variables on storage. Recognition experiments can be used to complement and extend recall investigations of variables influencing encoding processes.

REFERENCES

Battig, W. F., & Montague, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph*, 1969, **80** (3, Pt. 2).

Bower, G. H. A multicomponent theory of the memory trace. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation.* Vol. 1. New York: Academic Press, 1967.

D'Agostino, P. R. The blocked-random effect in recall and recognition. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 815–820.

EKSTRAND, B. R., WALLACE, W. P., & UNDERWOOD, B. J. A frequency theory of verbal-discrimination learning. *Psychological Review*, 1966, **73**, 566–578.

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

HOROWITZ, L. M., & MANELIS, L. Towards a theory of redintegrative memory: Adjective-noun phrases. In G. H. Bower & J. T. Spence (Eds.), *The psychology of learning and motivation.* Vol. 5. New York: Academic Press, 1972.

JACOBY, L. L. Effects of organization on recognition memory. *Journal of Experimental Psychology*, 1972, **92**, 325–331.

KINTSCH, W. Models for free recall and recognition. In D. A. Norman (Ed.), *Models of human memory.* New York: Academic Press, 1970.

LIGHT, L. L., & CARTER-SOBELL, L. Effects of changed semantic context on recognition memory. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 1–11.

MANDLER, G. Organization and memory. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation.* Vol. 1. New York: Academic Press, 1967.

MANDLER, G. Organization and recognition. In E. Tulving & W. Donaldson (Eds.), *Organization of memory.* New York: Academic Press, 1972.

MCCORMACK, P. D. Recognition memory: How complex a retrieval system? *Canadian Journal of Psychology*, 1972, **26**, 19–41.

MURDOCK, B. B., JR. Modality effects in short-term memory: Storage or retrieval? *Journal of Experimental Psychology*, 1968, **77**, 79–86.

POSTMAN, L. A pragmatic view of organization theory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory.* New York: Academic Press, 1972.

SHUELL, T. J. Clustering and organization in free recall. *Psychological Bulletin*, 1969, **72**, 353–374.

SLAMECKA, N. J. The question of associative growth in the learning of categorized material. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 324–332.

THORNDIKE, E. L., & LORGE, I. *The teacher's wordbook of 30,000 words.* New York: Columbia University, Teacher's College, Bureau of Publications, 1944.

TULVING, E. Theoretical issues in free recall. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory.* Englewood Cliffs, N. J.: Prentice-Hall, 1968. (a)

TULVING, E. When is recall higher than recognition? *Psychonomic Science*, 1968, **10**, 53–54. (b)

TULVING, E., & PEARLSTONE, Z. Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 381–391.

TULVING, E., & THOMSON, D. W. Retrieval processes in recognition memory: Effects of associative context. *Journal of Experimental Psychology*, 1971, **87**, 116–125.

UNDERWOOD, B. J. Attributes of memory. *Psychological Review*, 1969, **76**, 559–573.

UNDERWOOD, B. J., & FREUND, J. S. Word frequency and short-term recognition memory. *American Journal of Psychology*, 1970, **83**, 343–351.

WICKENS, D. D. Encoding categories of words: An empirical approach to meaning. *Psychological Review*, 1970, **77**, 1–15.

WINOGRAD, E., KARCHMER, M. A., & RUSSELL, I. S. Role of encoding unitization in cued recognition memory. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 199–206.

(Received September 9, 1972)