COGNITIVE PSYCHOLOGY 1, 86–98 (1970)

Free Recall of Trilingual Lists[1]

ENDEL TULVING AND VICTOR A. COLOTLA

University of Toronto, Toronto 5, Canada

Subjects fluent in English, French, and Spanish recalled lists consisting of words in one, two, or all three languages. The component of recall identified with primary memory was identical for unilingual, bilingual, and trilingual lists, but the secondary-memory component was appreciably greater for unilingual than multilingual lists. The main conclusions of the experiment were: (*a*) a multilingual person's different languages exist in relative isolation from each other; (*b*) organization of list words into higher-order memory units is more difficult between different languages than within a single language; and (*c*) lower recall of multilingual lists reflects reduced accessibility of information about list words.

For the past several years we have been doing free-recall experiments with bilingual subjects at Toronto. When we started this research, nothing had been published about bilingual free recall. Therefore, the first few experiments we did were undertaken to answer a very simple question: how well can a person fluent in two languages recall lists consisting of "unrelated" words in both languages, in comparison with recall from unilingual lists?

To the extent that the bilingual persons' two languages constitute separate coding systems (Kolers, 1963), it would not be unreasonable to expect that bilingual subjects (*S*s) might be able to recall more words from bilingual than from unilingual lists. Intralist interference may be less if the material consists of words in two languages; alternatively, recall may be greater from two stores whose combined storage capacity is likely to be greater than that of a single store.

The finding of equal recall from unilingual and bilingual lists would also make sense. We could assume that what is stored in a free-recall task are not specific words, but rather their meanings together with appropriate "language tags" permitting the reproduction of each word in its original language. If the storage of language tags occurs without any

[1] This research was supported by the National Research Council of Canada, Grant AP 39, and the National Science Foundation, Grant GB 3710. We thank Bennet B. Murdock, Jr., and Stephen A. Madigan for their help in carrying out this research.

"cost" to the storage system, recall should be the same from unilingual and bilingual lists.

Finally, as an eloquent testimony to the flexibility or immaturity of our discipline, it would be equally simple to rationalize the finding of lower recall from bilingual lists than from unilingual lists. For instance, we could assume that the language tags stored with word meanings do tax the storage capacity of the system, that is, storage of language tags can occur only at the expense of the storage of some of the word meanings. Experiments conducted by Stephen A. Madigan in our laboratory at Toronto have clearly shown that remembering certain types of additional information about list-words in a free-recall task does reduce recall of words as such. Even if each word is accompanied only by one bit of additional uncertainty—it appears on the left-hand or right-hand side of the screen, it is paired with a 1 or a 2, it is either the singular or plural form of a noun—and *S*s are asked to remember as many words as possible plus as much additional information as possible, recall of words is lower than recall of words under standard conditions. Hence it would not be too unrealistic to expect that the membership of list-words in one or the other to-be-remembered linguistic coding system constitutes a similar source of loss of storage capacity.

Since the time that we first thought these thoughts and conducted our initial experiments, a number of free-recall experiments with bilingual *S*s have been reported and our own original uncertainty as to the answer to our question should be reduced. But it is not. At least three sets of experimenters have reported that recall of bilingual lists of "unrelated" words is no different from recall of comparable unilingual lists (Kolers, 1965; Lambert, Ignatow, & Krauthamer, 1968; Nott & Lambert, 1968). Yet in at least six separate experiments of ours, we have observed a decrement in recall of bilingual lists in comparison with unilingual lists. In this paper we bring our results, at variance with those of other investigators, to the attention of those interested in memory, bilingualism, or both.

To report all of our experiments would produce a great deal of redundancy and would make for dull reading. We shall, therefore, describe only one experiment, our latest one and the most extensive. To magnify possible bilingual effects, we used speakers of English, French, and Spanish who had to recall unilingual, bilingual, and trilingual lists, but the essential findings reported here are highly similar to those that have emerged from our previous studies using English-Estonian as well as English-French bilinguals.

The experiment described here was designed for no profound theoretical purpose. We simply wanted to know how well *S*s could recall uni-

lingual, bilingual, and trilingual lists. As it frequently happens, however, the present results, too, have some theoretical implications, both for memory and for bilingualism. We shall consider these implications after we have described the method and results.

METHOD

Design

Three independent variables were combined orthogonally to produce 18 different experimental conditions. The *S*s were tested repeatedly under all 18 conditions. The independent variables were: (*a*) type of list—unilingual, bilingual, and trilingual; (*b*) length of list—12, 18, or 24 words; and (*c*) rate of visual presentation of the material—.5 sec/word and 2 sec/word.

The design will best be understood if we describe one "replication," since the total experiment consisted of six replications administered to each of the six *S*s. One replication corresponded to two experimental sessions in which the *S* was administered 54 lists, 27 lists in each session. The two sessions forming a replication differed only with respect to the rate of presentation of the material, fast (.5 sec/word) in one, and slow (2 sec/word) in the other. The 27 lists administered to the *S* in one session consisted of 9 lists of three different lengths: 12, 18, and 24 words. Within each level of length, three lists were unilingual (one each of English, French, and Spanish), three were bilingual (one each of English-French, English-Spanish, and French-Spanish), and three were trilingual. Each *S* was thus tested with a total of 324 lists in 12 sessions, that is, 18 different lists corresponding to each of the 18 different experimental conditions defined by the orthogonal combinations of three types of list, three list lengths, and two rates of presentation.

Materials

The pool of words used in the experiment consisted of 324 high-frequency English nouns and their French and Spanish translations, a total of 972 words. The 972 words were used to construct the 54 lists necessary for one replication. All lists were constructed with the aid of a computer program. Six different sets of 54 lists were constructed in all, one for each replication of the experimental design. Thus, the *S*s saw words from the same basic pool in successive replications, but lists were all different in different replications. The order in which the different sets of lists were presented to *S*s in successive replications was completely counterbalanced, as was the assignment of specific half-sets of 27 lists to fast and slow presentation conditions.

For unilingual lists words from the specified language were selected randomly from the pool without replacement. Bilingual and trilingual lists were also constructed by random selection of words without replacement, subject to the following two restrictions: (*a*) translation of any list-word in another language was not permitted in the same list; (*b*) the order of words in different languages was constructed to minimize the length of runs of words in the same language. Specifically, in bilingual lists one word from Language 1 and one word from Language 2 always occurred in serial positions 1 and 2, 3 and 4, 5 and 6, etc. The order of the two words occupying the pairs of adjacent input positions was random. In trilingual lists, one word from each of the three languages always occurred in a block of three successive input positions: 1, 2, and 3; 4, 5, and 6; 7, 8, and 9; etc. The order of the three languages within each block of three adjacent input positions again was random.

Thus, as the bilingual lists were constructed, the average probability with which a word in Language 1 was followed by another word in Language 1 was .25, while the average probability of its being followed by a word in Language 2 was .75. In trilingual lists, the average probability of a word in a given language being followed by another word in the same language was .11, while the average probability of it being followed by a word in either of the two other languages was .44.

Subjects

Four young men and two young women served as *S*s in the experiment. Three of them were university graduates, two were university students, and one was a final-year high school student. All six were proficient in English, French, and Spanish, and most spoke one or more other languages. All but one were born outside North America; but with the exception of one of the five, all had resided in Canada for at least six years. The *S*s received $1.50 per session and a $5.00 bonus at the completion of the experiment.

Procedure

Each *S* was tested individually in 13 sessions—a practice session and 12 experimental sessions. In the practice session the *S* was (*a*) administered a reading-speed test, consisting of three passages in each of the three languages, (*b*) provided with general information about the nature of the experiment and his task, and (*c*) given a practice test with nine lists—three unilingual, three bilingual, and three trilingual lists. The reading-speed test was administered in order to obtain an estimate of the *S*'s facility in the three languages, but since no obvious systematic relation

was observed between *S*'s performance in that test and in the experimental task proper, no further reference will be made to it.

In the experimental sessions, *S*s were told about the nature of each list before it was presented—whether it was unilingual, bilingual, or trilingual, and in the case of bilingual lists, the two languages involved—but they were not informed of the list length before each list. Their instructions throughout the experiment were to pay careful attention to all the words in each list and to recall as many as they could, in any order that the words occurred to them.

The words were presented by means of a closed-circuit TV system. The word lists generated by the computer were recorded by a TV camera and shown on the screen of a TV monitor, one word at a time, to the *S* sitting in an adjacent room. An electric timer was used to automatically control the rate of presentation.

At the end of each list *S* was allowed time for oral recall equivalent to twice the amount of time that had been taken to present the list. The *S*'s responses were recorded on a standard tape-recorder for subsequent transcription. The next list was presented approximately 10 sec after the end of the recall phase of the previous list. In addition, *S*s were given longer rest periods whenever they asked for them.

RESULTS

Free-recall performance, even on a single recall trial, can be readily analyzed into two independent components (e.g., Craik, 1968). The two components of recall are independent in the sense that certain experimental variables can be shown to have quite different effects on the two components. Following Waugh and Norman's terms (1965), we will refer to these two components as primary memory (PM) and secondary memory (SM).

In this experiment, each recalled word was classified as part of the PM or the SM component on the basis of the length of its intratrial retention interval (ITRI). The length of ITRI for any given item was specified in terms of the number of presentations and recalls of other items occurring between the presentation and recall of the given item. Thus, if an input sequence consists of items A, B, C, D, and E, and *S* recalls E, C, and D, in this order, ITRIs for these three items are 0, 3, and 3, respectively.

Words for which ITRI was 7 or less were classified as PM words, while those for which ITRI was 8 or more were classified as SM words. Thus, the number of words recalled on any trial is simply the sum of the number of PM and SM words. The selection of the ITRI of 7 as the defining criterion of PM is rather arbitrary, but two considerations are relevant:

(*a*) the mean number per trial of PM words thus defined was very close to 3.0, a conveniently round number that appears to be a reasonable estimate of the capacity of the PM component, and (*b*) this value was practically identical with an estimate of the capacity of PM based on the number of words recalled from the last four input positions, regardless of output position, in the present as well as in an earlier experiment (Tulving & Patterson, 1968).

PM and SM Recall

The grand mean of PM recall for the whole experiment was 2.97, the grand mean of SM recall, 3.33. The mean number of PM words was practically invariant for different levels of independent variables tested in this experiment: for unilingual, bilingual, and trilingual lists, for instance, the means were 2.98, 2.98, and 2.95, respectively. Only the rate of presentation had a somewhat more pronounced effect, the means being 2.90 for the fast, and 3.04 for the slow rate. Although the difference was small, analysis of variance showed it to be highly significant, $F(1,1615) = 20.38$, $p < .001$. No other main effects or interactions were found to be significant by the same analysis of PM recall.

All three main independent variables, however, had large and statistically highly significant effects on SM recall. Figure 1 presents a summary of these effects. In Figure 1 the mean number of SM words recalled is plotted against the total presentation time per list, in keeping with argu-

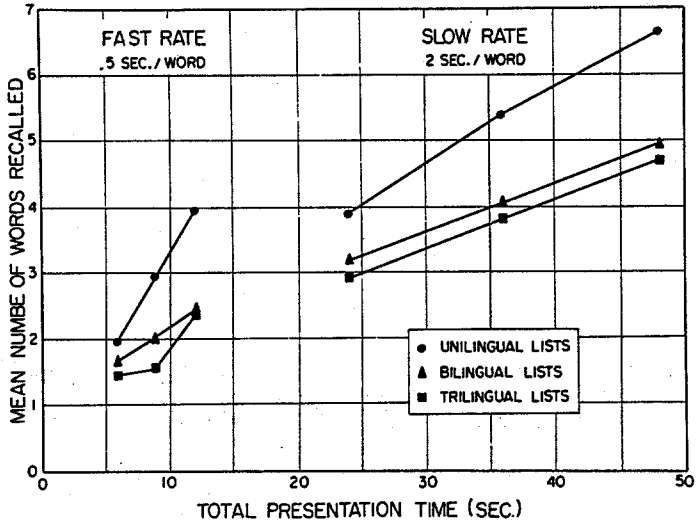

FIG. 1. Mean number of secondary-memory words recalled as a function of total presentation time and type of list.

ments advanced by Murdock (1960). The presentation time of a list is the product of rate of presentation per item and list length.

The most important finding depicted in Figure 1 is the inferior SM recall in bilingual and trilingual lists in comparison with unilingual lists. The three means, based on data pooled over all other conditions of the experiment, were 2.80, 3.05, and 4.13, for trilingual, bilingual, and unilingual lists, respectively. These means are approximately linearly related to the proportions of words in one language in the three types of list, namely 33%, 50%, and 100%.

Analysis of variance showed the List Type × Length interaction, as well as the List Type × Rate interaction to be significant, $F(4,1615) = 5.65$, $p < .001$, and $F(2,1615) = 3.04$, $p < .05$, respectively. The triple interaction among these variables yielded an $F < 1.0$. Apart from a significant interaction between list length and rate of presentation, $F(2,1615) = 13.09$, $p < .001$, no other significant interactions were found.

Recall of Words in Different Languages

Analysis of recall of words in different languages showed our sample of *S*s to be more efficient with English unilingual lists than with French and Spanish unilingual lists. The three means, for data pooled over list length and rate, were 7.67, 6.90, and 6.75, respectively. In recall of bilingual lists, English-French and English-Spanish were approximately equal, with means of 6.13 and 6.26, while French-Spanish was lower at 5.67. The mean number of words recalled from trilingual lists was 5.76.

A finding of some interest has to do with the relative frequency of recall of words in different languages depending upon the type of list in which they occurred. One might expect the superiority of one language over others to manifest itself in all types of list, or perhaps become even more pronounced in multilingual lists. Figure 2, however, shows that these expectations are not borne out by the data.

To obtain data shown in Figure 2 the "strength" of the three languages was rank-ordered for each *S* separately, depending upon his recall of unilingual lists. For the purpose of this analysis, each *S*'s data were again pooled over three list lengths and two rates of presentation. The language yielding highest recall under unilingual input conditions was regarded as the *S*'s "strong" language, while the language yielding lowest recall was regarded as his "weak" language.

The ordering of the three languages for unilingual lists in Figure 2 simply expresses a certain kind of imbalance of language skills among our trilingual *S*s. What is interesting, however, is the fact that words from *S*s' "weak" language (Spanish for four *S*s, French for two) thus defined are recalled better than the words from their two stronger languages both in

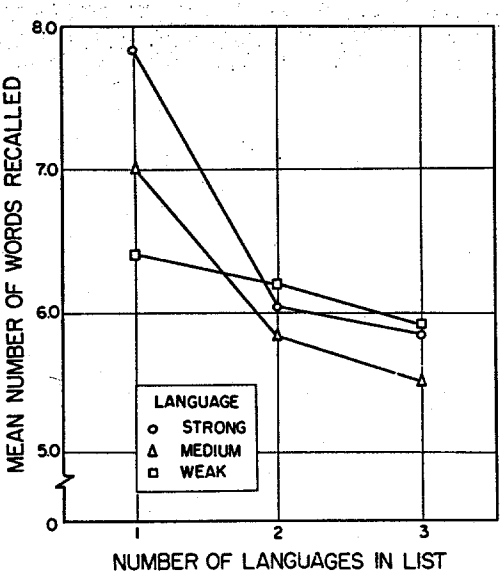

FIG. 2. Mean number of words recalled in different languages from unilingual, bilingual, and trilingual lists.

bilingual and trilingual lists. The impairment in recall of multilingual lists thus is produced to a large extent by impairment in recall of words in the language or languages recalled best in unilingual lists.

Recall Sequences in Multilingual Lists

The final finding worth mentioning has to do with the order in which words from different languages were recalled in bilingual and trilingual lists. One aspect of this order can be expressed in terms of conditional probabilities: recall of a word in Language 1 given that the immediately preceding word in recall was in Language 1, $P(R_1 \mid R_1)$, or in Language 2, $P(R_1 \mid R_2)$. To the extent that *S*s cluster recall by language, $P(R_1 \mid R_1)$ should be greater than $P(R_1 \mid R_2)$. Lack of a difference between these two conditional probability values, on the other hand, would demonstrate absence of any tendency to cluster by language. The data summarized in Table 1 show that neither of these outcomes occurred; instead, *S*s in this experiment showed "negative" clustering by language.

Table 1 presents conditional probabilities that a word in output position n, a word in one of the three languages, was followed by a word in any of the three languages in output position $n + 1$. Two conditional probability values are shown in each of the nine cells of Table 1: the upper one represents the data from trilingual lists, the lower one repre-

TABLE 1
Conditional Probabilities of Recall of a Word in a Particular Language in Output Position $n + 1$, Given Recall of a Word in Any One of the Three Languages in Output Position n (Upper figures—trilingual lists; lower figures—bilingual lists)

Language of word in output position n	Language of word in output position $n + 1$		
	English	French	Spanish
English	.211	.389	.400
	.239	.376	.385
French	.367	.234	.399
	.378	.287	.335
Spanish	.367	.378	.255
	.382	.326	.292

sents the comparable data from bilingual lists. In the latter case, conditional probabilities with which a word in a given language followed another word in the *same* language were calculated from the pooled data provided by two types of bilingual list. For instance, the probability that a French word followed another French word in bilingual lists (.287, according to Table 1) was based on data from English-French and French-Spanish bilingual lists. In all cases the conditional probability values are proportions of approximately one thousand observations.

The pattern of data in Table 1 makes it quite clear that *S*s, having recalled a word in one language, more often than not proceeded to recall the next word in a different language. No significance tests were performed on these data. Given the large number of observations on which the data are based and the reliability of the overall pattern, the facts seem to be firm. In addition, every single *S* exhibited this pattern, and neither the length of list nor the rate of presentation seemed to have any systematic effect on this phenomenon of "negative" language clustering.

DISCUSSION

The principal findings of this experiment were the following: (*a*) recall of words from primary memory was identical for unilingual and multilingual lists; (*b*) recall of words from secondary memory was greater in unilingual than bilingual lists, and greater in bilingual than trilingual lists; (*c*) recall of words in the language producing best recall under uni-

lingual conditions was greatly impaired under bilingual and trilingual conditions, while recall of words in the language producing least unilingual recall suffered relatively little impairment in bilingual and trilingual lists; and (*d*) in recalling words from bilingual and trilingual lists *S*s tended to follow the recall of a word in one language more frequently with the recall of a word in the other or another language than with a word in the same language.

Absence of any differences in primary memory, in the presence of sizable differences in secondary memory, is not particularly surprising at the empirical level: other variables are known to exist that have no effect on the primary-memory component of free recall, but do have an effect on the secondary-memory component. The present results simply extend the range of such variables to include the linguistic composition of lists. They also complement the general conclusions reached by Kintsch and Kintsch (1969) that semantic relationships between words in different languages influence tasks involving secondary memory, but not those involving primary memory. Since the primary findings of interest in the present experiment have to do with the secondary-memory component of recall, in what follows we will largely ignore the observed stability, across experimental conditions, of the primary-memory component—although this finding has interesting theoretical implications.

Why are multilingual lists recalled less efficiently than unilingual lists? Could it be that the storage of relevant information is impaired under multilingual input conditions? One might assume, for instance, that in studying multilingual lists *S*s spend a certain amount of time for the identification of the language in which a given group of letters spells a meaningful word. This time would then not be available for appropriate storage of words to be remembered, and recall of multilingual material would be impaired because of its inadequate initial storage.

Our data do not provide much support for the tenability of the hypothesis of impaired storage of multilingual lists. First, if we assume that the primary-memory component of recall represents a relatively "pure" measure of what is stored, then the uniformly identical PM recall measures for unilingual and multilingual lists contraindicate differential storage. Secondly, and perhaps more importantly, if storage of multilingual lists had been less efficient than storage of unilingual lists, because of time consumed by switching from one language to another, larger differences between unilingual and multilingual lists should have occurred under the conditions of fast rate of presentation than the slow rate. The results showed that the impairment in recall of bilingual and trilingual lists, if anything, was greater under the slow rate than under the fast rate of presentation.

Since we cannot detect much evidence in our data pointing to differences in storage of unilingual and multilingual lists, we are inclined to interpret the differences in recall of different types of list in terms of the differential facility with which words from different types of list can be retrieved: we think that differences in secondary-memory recall between unilingual and multilingual lists reflect differences in accessibility of otherwise equally available stored information (Tulving & Pearlstone, 1966).

It is not known what determines the accessibility of available information under the conditions of the free-recall task. Specification of relevant retrieval cues in such a situation constitutes one of the truly important objectives of research in memory. All we can do at the present time is to assume that information about at least some items in the memory store somehow becomes accessible through other items, however recalled, acting as retrieval cues for this information. If two items, in the course of presentation of the list, have become organized into a higher-order unit of information, recall of one will bring about the recall of the other. To the extent that such organization of individual list-items into higher-order units is difficult, accessibility of stored information is impaired. The lowered recall of words from multilingual lists thus is regarded as a consequence of impaired organization of words across language boundaries, or impaired organization of words within a language owing to the special construction of multilingual lists used in the present experiment. As the reader may remember, no more than two words from the same language ever occupied adjacent input positions in bilingual and trilingual lists, and even then the frequency of the unilingual bigrams was low.

This analysis is consistent with the observation of "negative" language clustering which demonstrated two things: (*a*) language clustering by *S*s to an extent considerably greater than that found in the input lists, and (*b*) language clustering by *S*s to an extent considerably less than would have been expected by chance as, for instance, defined by a series of unbiased coin tosses. The *S*s in this experiment were apparently resisting the deleterious effects of list structures that destroyed the temporal contiguity of words in one language in multilingual lists, but they were not completely successful in overcoming these effects.

The analysis is further consistent with the finding that the impairment in recall of multilingual lists tended to vary directly with the amount of material recalled under unilingual conditions. If variations in recall of number of words in unilingual lists are, at least to some extent, a consequence of unitization of words into higher-order units, and if the structure of multilingual lists we used makes it difficult to organize words within a given language system, one would expect greater impairment under

conditions where greater degrees of organization are required for "normal" recall, that is, recall under unilingual input conditions. Lambert *et al.* (1968) as well as Nott and Lambert (1968) have also reported impairment in recall of bilingual lists in comparison with readily organizable unilingual lists.

To summarize, we have argued that the differences in the secondary-memory component of recall between unilingual and multilingual lists are not attributable to differences in the storage of relevant information, but rather to the lowered accessibility of words from multilingual lists owing to the impaired organization of individual list items into higher-order units. The particular structure of bilingual and trilingual lists we used, we believe, interfered with optimal organization and was thus directly responsible for the results we obtained. We would expect that under conditions where retrieval of list words is less dependent upon intralist organization, such as in cued recall and recognition tests, or under conditions where within-language organization is less affected by the linguistic composition of the list, as in the case of blocked presentation of words in different languages, impairment in retrieval of bilingual and trilingual lists would be much smaller in extent, if it exists at all.

To the extent that the explanation of our primary findings makes sense, the assumptions on which the explanation is based may also be worth examining more closely. We find particularly intriguing the assumption that the two languages of a bilingual person, or three languages of a trilingual person, normally exist in relative isolation from one another (Kolers, 1963). It is intriguing because it suggests the possibility that the isomorphic relation between any two languages of a multilingual speaker exists only to the extent that it is imposed on the two languages by the application of the rules of translation, and that the relation does not exist simply by virtue of the person's ability to speak and comprehend the two languages. Such ability, as we know, can exist independently of any formal knowledge of grammar, while the skill of translation is critically dependent upon such knowledge.

Finally, we have no speculations to offer about the finding that impairment of recall of multilingual lists is produced to a large extent by impairment of recall of words in the language that is recalled best in unilingual lists. We do not believe that it is simply a matter of regression toward the mean. We have observed the same phenomenon in most other bilingual free-recall experiments we have done, and Nott and Lambert (1968) have reported a similar finding. The phenomenon appears to be real and intriguing, but the deciphering of the message it contains has to await further progress in our understanding of free recall, multilingual minds, or both.

REFERENCES

CRAIK, F. I. M. Two components in free recall. *Journal of Verbal Learning and Verbal Behavior,* 1968, **7,** 997–1004.

KINTSCH, W., & KINTSCH, E. Interlingual interference and memory processes. *Journal of Verbal Learning and Verbal Behavior,* 1969, **8,** 16–19.

KOLERS, P. A. Interlingual word associations. *Journal of Verbal Learning and Verbal Behavior,* 1963, **2,** 291–300.

KOLERS, P. A. Bilingualism and bicodalism. *Language and Speech,* 1965, **8,** 122–126.

LAMBERT, W. E., IGNATOW, M., & KRAUTHAMER, M. Bilingual organization in free recall. *Journal of Verbal Learning and Verbal Behavior,* 1968, **7,** 207–214.

MURDOCK, B. B., JR. The immediate retention of unrelated words. *Journal of Experimental Psychology,* 1960, **60,** 222–234.

NOTT, C. R., & LAMBERT, W. E. Free recall of bilinguals. *Journal of Verbal Learning and Verbal Behavior,* 1968, **7,** 1065–1071.

TULVING, E., & PATTERSON, R. D. Functional units and retrieval processes in free recall. *Journal of Experimental Psychology,* 1968, **77,** 239–248.

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

WAUGH, N. C., & NORMAN, D. A. Primary memory. *Psychological Review,* 1965, **72,** 89–104.

(Accepted August 27, 1969)