The Effect of Sentence Context on the Stability of Phonemic and Semantic Memory Dimensions[1]

DONALD A. WALTER[2]

University of Michigan, Ann Arbor, Michigan 48104

Phonemic and semantic memory dimensions of words presented in sentence or non-sentence context are examined with an eye toward evaluating the relationship between multistage models of memory and sentence word encoding. The former implicitly assume that short-term memories contain phonemic information while long-term memories contain semantic information. Results from a modified probe recognition task, in which the sound or meaning of a word is the unit of recognition, show that semantic information dissipates far more rapidly than phonemic, except within a sentence context. This is inconsistent with a hypothesis of semantic dominance of long-term memory.

The experiment to be reported here deals with the effects of sentence and nonsentence context on the phonemic and semantic dimensions of word memory. In doing so it also seeks to evaluate the claim that long-term stages of sequential multistage memory models contain information of a semantic nature and that short-term stages are limited to primitive phonemic information.

The great majority of recent studies involving language behavior have drawn heavily from two separate sources, traditional linguistics and verbal learning and memory. The former has served as a source of theories while the latter has provided the techniques for their evaluation. This has proven to be an unfortunate situation, for the preoccupation of linguistics with grammatical variables has inhibited language research in two ways. First, a great deal of systematized knowledge about verbal learning and memory has been somewhat ignored. Second, the magnitude of effects of grammatical manipulations has been overrated, for recent studies have shown them to be overwhelmed by nongrammatical manipulations. Sachs (1967), for example, has shown that semantic changes in sentences are far easier to detect than grammatical changes, and Paivio (1971) has found that recall differences attributed to grammatical manipulations (subject and object nominalizations) by Rohrman (1968) and Rohrman and Polzella (1968) were actually due to noun imagery differences.

For these reasons I have redirected my efforts toward assessing sentence encoding behavior in terms of learning and memory models. Specifically, using models of word memory as a starting point, I will assess the effects of sentence context on the phonemic and semantic memory dimensions of words and their relationship to multistage models of memory. First, however, some discussion of memory models is in order.

Information-processing models of Sperling (1967, 1968) and Norman and Rummelhart (1970) posit a multilevel encoding process in which the form of the memory representation changes at each level. The primary levels in such models hold information of a primitive nature, normally iconic information of a

[2] Now at the University of Notre Dame, Notre Dame, Indiana 46556.

phonemic or visual nature. The same general representations are also used in the familiar two-stage memory models of Atkinson and Shiffrin (1968) and Waugh and Norman (1965). They assert two separate memories, a short-term memory and a long-term memory. Although it is implied in such models that short-term memory holds primarily phonemic information and long-term memory holds primarily semantic information, the only strong statements to this effect are those of Sperling. Because of this, a great deal of research has surrounded the identification of memory trace characteristics at each level, especially in terms of phonemic and semantic representations. A review of such studies may be found in Shulman (1971), but it would be worthwhile to mention some of the major findings here.

Conrad, Baddeley, and Hull (1966), Conrad (1967) and Wickelgren (1965) found in separate studies that the manipulation of acoustic similarity in a short list of letters had deleterious effects on the immediate recall of such lists, while semantic and graphic similarity had relatively small effects. In addition, Wickelgren (1966) and Conrad (1964) found that intrusions in such recall protocols were acoustically or phonemically related to the presented letters. Kintsch and Buschke (1969) used a sequential probe recognition technique to determine the retention interval at which an acoustic or semantic similarity between words in a list would have the greatest effect. With acoustic similarity, strong effects were found at small retention intervals; and with semantic similarity, at large intervals.

Shulman (1970) used a modified probe recognition procedure capable of assessing not only the strength of the word memory, but the strength of the phonemic and semantic attributes of the memory trace. In the ordinary probe recognition experiment, a list of words is presented to a subject and followed by a single, isolated word. The subject simply indicates whether the single word was a member of the list. The major reason for the adoption of this method in word memory studies is that output interference is virtually eliminated. Since only one word memory retrieval is required, the effects of retrieval of other words in the list presumably cannot interfere. Shulman presented subjects with a list of 10 words followed by a special probe recognition test. If he wanted the subject to make an ordinary recognition of the probe word, an *I*, standing for identical, preceded the test. If the subject was to make a recognition of the semantic properties of the word, an *S*, standing for synonym, was shown. For the phonemic dimension, an *H*, standing for homonym, served as the cue. Hence, subjects were making recognitions either on the basis of all available memory dimensions, on the basis of meaning only, or on the basis of the sound of the word. Recognition scores for the three tasks were then used to determine the saliency of the corresponding dimensions of the memory trace. This modified probe method is capable of an exact examination of the makeup of the memory trace. For this reason the Shulman procedure was used in the research presented here.

My concern is to determine not only how the memory of a word changes when it is part of a sentence, but through the use of postlist retention intervals to investigate how this change is related to traditional representations of short- and long-term memories. Common sense would dictate that a word in a sentence would eventually be encoded in terms of its long-term semantic dimensions. However, it is questionable whether sentences go through a primitive phonemic encoding stage prior to this, for there is evidence that higher order components like words and sentences are perceived faster than their subcomponents (Savin & Bever, 1970). Contact with a long-term semantic memory may be immediate.

METHOD

Subjects

Seventy-two University of Michigan female undergraduates served as subjects. They were each paid

$1.75 for their participation in the experiment; performance bonus was additional. Each subject served for approximately 1 hour and was randomly assigned to a particular counterbalancing condition.

Experimental Design

The basic experiment consisted of 72 probe recognition trials. The independent variables were word presentation context, retention interval and dimension of the recognition test. These were employed in a completely within-subjects design. Presentation context was varied by presenting the to-be-recognized word in an intact sentence, or in a word string obtained by scrambling the sentence words surrounding the to-be-recognized word in a quasi-random fashion so as to eliminate meaningful sequences. Retention intervals of 5 and 20 seconds were filled with mathematics problems with a financial reward for their correct solution. Three recognition tasks were used, one based on recognition of phonemic identity, one on semantic identity, and the other on total identity of a probe word with a word in the presentation string.

Catch trials, on which "no" was the correct answer to the probe recognition test, comprised one-third of the recognition tests. Thus, out of the six trials for each combination of the three independent variables, two were catch trials.

Two orders of the presentation of conditions were randomly selected. Each was used for one-half of the 72 subjects. The counterbalancing of materials within conditions was complete. Over the 72 subjects, each word string appeared in each combination of retention interval, context and recognition test.

Procedure

Each subject was seated in front of a cathode-ray-tube display device. After he read the instructions he was given six practice trials and 72 experimental trials of the following type: A ready signal appeared on the screen and was followed in 1 second by the whole presentation of a 12-word string, either a sentence or a scrambled word string. The subject read this string aloud at a 2-second rate, paced by a metronome quietly plinking in the background. The presentation time was such that the entire string disappeared 1 second after the twelfth plink. This gave the subject a starting lag of 1 second to insure he read the entire string. A retention interval of 5 or 20 seconds followed the string and was filled with arithmetic problems of the type $(x \times y) - z$, which appeared at the rate of one every 2.3 seconds. The subject was paid 0.5 cent for each problem he solved correctly. He said his answer aloud over an intercom.

After the last problem in the retention interval, another ready signal appeared on the screen. This signalled the subject to stop doing the arithmetic problems and to prepare for a recognition test. One second

later a large *H*, *S*, or *I* appeared on the screen. The subject had been taught previously that these stand for homonym, synonym and identical, respectively. This letter, the probe cue, indicated to the subject which of three types of recognition he was to make of the probe word, which appeared to the right of the probe cue 0.5 second later. The subject had simply to decide whether the probe word had the cued relationship to any of the words in the preceding word string. He responded by pressing either a "yes" or "no" key in front of him. In addition, after his response, he pressed one of three keys to indicate his response confidence as "probably," "sure" or "very sure."

The subject was told that the incidence of correct "no" answers would be one-third. Such information tends to stabilize response bias (see Parks, 1966). In addition, since latencies for the "yes" and "no" responses were recorded, the subject was told to respond not only with his first impulse, but to guess if 4 seconds had passed since the appearance of the probe word (judgment of this interval was left to the subject). This was done to ensure that the latency distribution would not be overly skewed and to raise the error rate in the easier conditions. Five seconds after he made his confidence response, the start of the next trial was signalled by another ready signal.

Apparatus

A cathode-ray-tube display coupled to a PDP-1 computer was used to present the stimuli and order the conditions. Responses and their latencies were automatically recorded on paper tape with a list of the experimental conditions. The subject was isolated in a dimly lit room. After the six practice sentences, only verbal contact was maintained. An intercom was used to assure that the subject was reading the word strings properly and to answer any questions he might have during the experiment.

Materials

Seventy-two words were selected with one restriction in mind, that they have both unambiguous synonyms and homonyms. These varied widely in word class and frequency in usage. Of these words, called critical words, 48 were nouns and 24 were adjectives or verbs. For each word a 12-word sentence was constructed using words of the same general frequency to hide the identity of the critical word. The sentences were constructed so that six nouns and three adjectives or verbs occurred in each of the middle eight serial positions.

Each sentence was scrambled randomly to produce an ungrammatical string of words. If the scrambling resulted in a meaningful combination of adjacent words, words were displaced to eliminate this. Thus, the sentence *The thief cares little about causing the grim pain his victims suffer* was scrambled to produce the word

string *Grim his little thief suffers cares the about pain causing victims the*. The critical word *pain* retained its serial position in the scrambled version. Although each critical word was used in only one serial position, the Thorndike-Lorge frequency of occurrence was equated across positions.

For each of the 72 critical words, catch words were selected that duplicated the appropriate homonym or synonym in all respects except their relationship to the critical word. So, for example, the critical word *pain* had a synonym *hurt* and a homonym *pane*. For catch trials in the homonym recognition task the word *core* would appear instead of the word *pane*. The former has the same number of syllables, belongs to the same word class, and is a homonym itself. This last requirement was important, otherwise the subject could have responded only to the homonymity of the probe word without making any contact with his memory for the critical word. The same type of restrictions were made for catch words in the case of synonym tests. (These words were not, of course, synonymous with any word in the presentation string.) Since the characteristics of the identical probe words (the critical words) and their synonyms were quite similar, the same catch words were used for both tests so that a comparison of latencies could be made across the two recognition tests. In the above case the synonym *hurt* had as its catch word *dust*, which was also the catch word for the identical probe word, the critical word *pain*. Notice that *dust* duplicates the dual use of *hurt*, which could be used as a verb or a noun in the appropriate context. All efforts were made to duplicate the probe words as exactly as possible.

The arithmetic problems that filled the retention interval were drawn at random, excepting easy combinations involving computation with fives and ones. In such lists some duplications were inevitable since the number of problems presented during a typical session was over four hundred.

RESULTS

There are three dependent variables of interest: correct recognition, confidence ratings and response latencies. The first of these may be broken down into the hit rate (the proportion of correct yes responses) and the false positive rate (the proportion of "yes" responses to catch words or incorrect recognitions of new words as old). These scores are shown in Table 1.

The hit proportions were the source of the analysis of variance. All three main effects were highly significant, as was the interaction between context and recognition task: for context, $F(1, 11) = 94.6$, $p < 0.001$; for retention interval, $F(1, 71) = 15.9$, $p < 0.001$; for recognition task, $F(2, 142) = 3.81$, $p < 0.025$; and for the interaction of between context and recognition task, $F(2, 142) = 6.84$, $p < 0.005$. The remaining interactions were not statistically significant.

One of the problems in using hit proportions as an estimate of strength is that response bias differences reflected by false positive

TABLE 1

HIT AND FALSE POSITIVE RATES FOR TWO RETENTION INTERVALS
AND THREE RECOGNITION TASKS

Statistic	Retention interval	Context	Task type		
			Homonym	Synonym	Identical
Proportion hits		Sentence	.868	.888	.955
	5	Scrambled	.816	.760	.795
		Sentence	.855	.885	.918
	20	Scrambled	.757	.680	.743
Proportion false positives		Sentence	.082	.097	.010
	5	Scrambled	.153	.174	.069
		Sentence	.174	.153	.062
	20	Scrambled	.222	.330	.097

rates must be assumed to be negligible. This was not the case in these three recognition tasks. A recognition score that corrects for such response bias is the high-threshold correction for guessing (Luce, 1963). This involves application of the formula

$$\frac{\text{Hit Rate} - \text{False Positive Rate}}{1 - \text{False Positive Rate}}$$

The corrected recognition scores derived from application of this formula are shown in Figure 1.
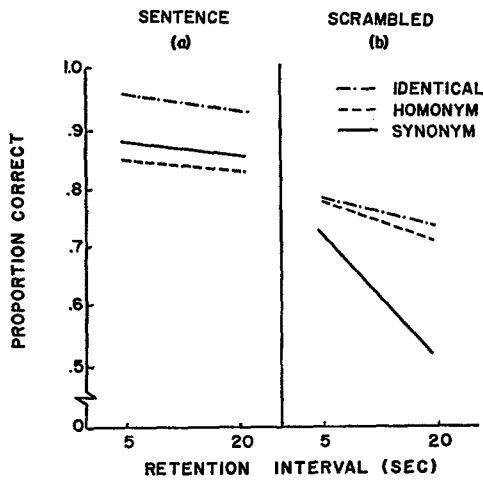

FIG. 1. Proportion correct recognition (corrected for guessing) for three recognition tasks as a function of presentation context and retention interval.

As is readily obvious from Figure 1, correct recognition benefited from sentence context for all three tasks and decreased as a function of retention interval. There is also a general superiority of identical recognition over both homonym and synonym recognition. The rank ordering of homonym and synonym recognition scores, however, was influenced by context. In Figure 1b, where the critical word is presented as part of a scrambled string, not only is homonym recognition superior, but synonym recognition decreases rapidly as a function of retention interval. However, when the critical word is presented as part of a sentence, synonym recognition exceeds that of homonym, and the rapid loss in the former disappears. In addition, sentence context slightly reduces the forgetting rate for both identical and homonym recognition.

The collection of confidence ratings was originally intended to generate points for a signal-detection analysis. Because of the very low false positive rates for identical recognition this analysis was impossible. It serves no purpose to discuss the mean confidence ratings as a function of the independent variables, since they duplicated the trends in Figure 1 exactly.

Even though all efforts were made in instructing subjects to respond soon after the probe or catch word appeared, some were unable to do so. There were no very long latencies, but the overall distribution of latencies was skewed. Because of this, a log transformation was applied to each latency prior to the analysis of variance. All latency means reported above are actually antilogs of the transformed score means. However, they vary little from their respective medians and are thereby representative of central tendencies.

Shown in Table 2 are mean latencies for correct probe- and catch-word responses, collapsed over rentention interval. Separate analyses of variance were done for each. For probe words, significant main effects were found for context $F(1, 71) = 39.7, p < 0.001$, recognition task, $F(2, 142) = 101.8, p < 0.001$, and retention interval, $F(1, 71) = 6.2, p < 0.025$. Of the interactions, only that between retention interval and recognition task was significant, $F(2, 142) = 4.42, p < 0.025$.

The analysis for catch-word latencies showed much the same pattern. The significant main effects were presentation context, $F(1, 71) = 9.4, p < 0.005$, retention interval, $F(1, 71) = 26.3, p < 0.001$, and recognition task, $F(2, 142) = 113.4, p < 0.001$. Two interactions were significant, those between context and recognition task, $F(2, 142) = 8.5, p < 0.001$, and between context and retention interval, $F(1, 71) = 7.0, p < 0.025$. Table 2 shows a

TABLE 2

MEAN REACTION TIME (IN MILLISECONDS) FOR CORRECT "YES" RESPONSES AND CORRECT
REJECTIONS AS A FUNCTION OF PRESENTATION CONTEXT AND RECOGNITION TASK

Response	Context	Task type		
		Homonym	Synonym	Identical
Correct "yes" responses	Sentence	2233	2406	1769
	Scrambled	2428	2631	1846
Correct rejections	Sentence	3576	3016	2209
	Scrambled	3054	3066	2046

general reduction of response latencies for probe words with respect to catch words.

In general, identical recognition (both probe and catch words) was fastest, homonym second, and synonym slowest. However, there is a glaring exception to this in the case of correct rejections of homonym catch words. While sentence context causes a general reduction of probe-word latencies for all three recognition tasks, the opposite occurs for catch-word latencies in the case of both identical and homonym latencies, with a dramatic increase in the case of the latter.

DISCUSSION

The questions to which this experiment was directed were as follows. How does the inclusion of a word in a sentence influence the memory dimensions of that word? More specifically, what are the relative changes in phonemic and semantic dimensions of the memory trace, and in what ways do these changes reflect multistage memory encoding processes?

It is clear that the presentation of a word in a sentence has a dramatic effect on the semantic dimension of memory. The interaction of presentation context and recognition test shown in Figure 1 attests to this. There is a reversal of homonym and synonym recognition as context is changed, as well as a radical reduction in synonym forgetting. It is also true that sent-

ence context has a general facilitative effect on all three recognition tests, more so for identical and synonym than homonym. On the other hand, even though there is an apparent general tendency for sentence context to reduce forgetting rates, this tendency is not significant.

As for the properties of phonemic and semantic memory dimensions, this study replicates Shulman's (1971) finding that phonemic dimensions dominate in noncontextual word memory. However, the results are inconsistent with a phonemic–semantic two-stage model of any type since homonym recognition scores decrease much more slowly than synonym scores during the retention interval. These results are also in apparent conflict with those of Kintsch and Buschke (1969), who found that manipulation of semantic and acoustic intra-list interference had differential effects on the immediate probe recall of words occurring early and late in the list, respectively. Unfortunately, a direct comparison of the studies is uninformative not only because of the obvious procedural differences, but because different ranges of retention intervals are in question.

Although semantic dimensions are superior to phonemic dimensions in sentence context, whether there is an actual reduction in phonemic memory from that of nonsentence context is in doubt because performance was superior overall. However, if identical recognition performance is used as a baseline for general improvement in this condition, there is

a relative reduction in homonym recognition in sentence context. This interpretation is even more compelling because of the possibility of a ceiling effect for identical recognition, which underestimates the baseline in sentence context.

This sentence context superiority of semantic dimensions over phonemic does not appear, however, when response latencies are considered, except in the case of catch words. In Table 2, consider only the latencies for correct recognitions of probe words. In all cases except synonym recognition in sentence context, probe-word latencies complement recognition scores. There is a general effect of sentence context, recognition task, and retention interval (not shown), but no interaction of recognition task and context. In both context conditions, identical recognition was most rapid, followed by homonym and synonym in that order. This indicates that the effects of sentence context are not to provide a quicker route to the semantic dimension of the word memory, but possibly to elaborate the number of routes, the selection of which may take some time. Thus, it is easier to decide whether adequate semantic information is there, but it takes longer when the context is sentential.

Evidence that an additional process may be occurring can be seen when latencies for catch words are considered. The reduction of latencies as a function of sentence context occurs only for semantic recognition (50 milliseconds), while identical latencies show a relatively small increase (163 milliseconds), and phonemic latencies show an enormous increase (522 milliseconds). The same latency difference shows up in correct recognitions of probe words (misses), but due to the small number of data points the data are rather unreliable. Evidently the subject is engaging in some elaborate activity in order to reject homonym catch words, activity that is apparent in no other condition. A likely possibility is that some decisions are being made by reconstructing phonemic dimensions from other dimensions of information. This is interesting

in itself, but also indicates that homonym recognition is an even greater overestimate of phonemic dimension saliency because the subject is making accurate rejections of homonym catch words by using other memory dimensions. This and earlier evidence is a strong indication that phonemic dimensions are little used in sentence context word encoding.

There are two viable explanations of the reversal of synonym and homonym recognition as a function of sentence context. The first considers an intentional strategy shift by the subject as a function of context. Perhaps the subject finds it more practical to subvocally rehearse in the case of word-string presentation, but to attempt to grasp the meaning of the sentence in the sentence context case. This is unlikely, for two studies have found subjects unable to selectively encode words in terms of either phonemic or semantic memory dimensions (Cermack, Schnorr, Buschke & Atkinson, 1970; Buschke & Lenon, 1969).

The more viable alternative concerns the steps in sentence perception, at least within the confines of this task. It is tautological that phomene perception is necessary to segment an incoming acoustic signal into morphemes. The major function of phonemic memory is probably that of a temporary iconic repository awaiting extraction of semantic information. However, in the case of written material visually presented, no such segmentation is needed. Thus, no temporary phonemic stage is necessary. Furthermore, there is no indication that phonemic dimensions are temporary at all, for both homonym and synonym recognition remain equally stable over the retention interval (Figure 1a). It seems that they reflect a relatively permanent type of memory just as stable as semantic memory but little used in sentence encoding.

In summary, a special probe recognition technique was employed to evaluate the contribution of phonemic and semantic memory dimensions to word memory in sentence and nonsentence presentation contexts. In addition, the validity of sequential multistage

memory models using these dimensions was assessed. Phonemic dimensions dominated semantic dimensions in nonsentential presentation context even in terms of stability over a retention interval. Semantic memory dimensions were found to be rather fragile over a short 20-second retention interval, but only in nonsentence context. This evidence is inconsistent with conceptions of memory as a two-or-more memory system in which the earlier stages are phonemic and the later stages are semantic. Such models predict rapidly decaying phonemic dimensions and relatively stable semantic dimensions. The opposite was found to be the case, and the introduction of sentence context was necessary to arrest the rapid decay of semantic memory dimensions.

The similarity of retention functions for homonym and synonym recognition in the sentence context presentation condition was interpreted as evidence for a relatively permanent phonemic memory that is little used in sentence memory but may act as a single morphological unit.

REFERENCES

ATKINSON, R. D., & SHIFFRIN, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation.* New York: Academic Press, 1968. Vol. 2.

BUSCHKE, H., & LENON, R. Encoding homophones and synonyms for verbal discrimination and recognition. *Psychonomic Science,* 1969, **14**, 269–270.

CONRAD, R. Acoustic confusions in immediate memory. *British Journal of Psychology,* 1964, **55**, 75–84.

CONRAD, R. Interference or decay over short retention intervals? *Journal of Verbal Learning and Verbal Behavior,* 1967, **6**, 49–54.

CONRAD, R., BADDELEY, A. D. & HULL, A. J. Rate of presentation and the acoustic similarity effect in short-term memory. *Psychonomic Science,* 1966, **5**, 233–234.

CERMACK, G., SCHNORR, J., BUSCHKE, H., & ATKINSON, R. C. Recognition memory as influenced by differential attention to semantic and acoustic properties of words. *Psychonomic Science,* 1970, **19**, 79–81.

KINTSCH, W., & BUSCHKE, H. Homophones and synonyms in short-term memory. *Journal of Experimental Psychology,* 1969, **80**, 403–407.

LUCE, R. D. A threshold theory for simple detection experiments. *Psychological Review,* 1963, **70**, 61–79.

NORMAN, D. A., & RUMELHART, D. E. A system for perception and memory. In D. A. Norman (Ed.), *Models of human memory.* New York: Academic Press, 1970.

PAIVIO, A. Imagery and deep structure in the recall of English nominalizations. *Journal of Verbal Learning and Verbal Behavior,* 1971, **10**, 1–12.

PARKS, T. F. Signal-detectability theory of recognition-memory performance. *Psychological Review,* 1966, **73**, 44–58.

ROHRMAN, N. L. The role of syntactic structure in the recall of English nominalizations. *Journal of Verbal Learning and Verbal Behavior,* 1968, **7**, 904–912.

ROHRMAN, N. L., & POLZELLA, D. J. Recall of subject nominalization. *Psychonomic Science,* 1968, **12**, 376–377.

SACHS, J. Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics,* 1967, **2**, 437–442.

SAVIN, H. B., & BEVER, T. G. The nonperceptual reality of the phoneme. *Journal of Verbal Learning and Verbal Behavior,* 1970, **9**, 295–302.

SHULMAN, H. Encoding and retention of semantic and phonemic information in short-term memory. *Journal of Verbal Learning and Verbal Behavior,* 1970, **9**, 499–508.

SHULMAN, H. Similarity effects in short-term memory. *Psychological Bulletin,* 1971, **75**, 399–415.

SPERLING, G. Successive approximation to a model for short-term memory. *Acta Psychologica,* 1967, **27**, 282–292.

SPERLING, G. Phonemic models for short-term auditory memory. *Proceedings of the American Psychological Association,* 1968, **4**, 63–64.

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

WICKELGREN, W. A. Short-term memory for phonemically similar lists. *American Journal of Psychology,* 1965, **78**, 567–574.

WICKELGREN, W. A. Phonemic similarity and interference in short-term memory for single letters. *Journal of Experimental Psychology,* 1966, **71**, 396–404.