JUDGEMENTS OF FREQUENCY WITH INCIDENTAL LEARNING

ROBERT JAMES ROSE

JUDGMENTS OF FREQUENCY WITH INCIDENTAL LEARNING

Robert James Rose, B.A., B.Sc.

A Thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science

Department of Psychology
Memorial University of Newfoundland

August 1974

ABSTRACT

Subjects were given a long list of words and required to carry out one of three tasks: (1) evaluate the "goodness" implied by a word; (2) code each word arithmetically according to the arbitrary values assigned to the vowels a, e, i, o, u, by the experimenter; (3) find a rhyme for each word. Within each list, the experimental words occurred 2, 3, or 5 times with spacings of 0, 1, 2, 4, 8, or 16 items between repetitions. In an unexpected judgment of frequency test given at the end of the initial task, it was found that the evaluation group had the highest mean judgment and was most accurate while the coding group had the lowest mean judgment. The rhyming group was not significantly different from the evaluation group on mean judgment of frequency nor from the coding group on accuracy. All double interactions among the 3 main factors were significant in the analysis of the judgment of frequency results. The effects of both presented frequency and spacing were least pronounced in the coding condition, while the frequency x spacing interaction was due to a more pronounced spacing effect at the frequency of 5 than at the two lower levels of frequency.

The three groups differed among themselves in number of recognition errors, with the evaluation group making fewest errors and the coding group making most. The number of correct recognitions made by a subject correlated significantly with his judgment of frequency. The results require

of any model for the memorial representation of event frequency that it include the effects of real frequency, level of processing, spacing, and the interactions among these. It is suggested that such a model would be of a multiple-process type with a concept of "derived strength" involved in the retrieval of information concerning event frequency and a concept of "list markers" involved in the effect of spacing.

ACKNOWLEDGEMENTS

The author wishes to thank Ted Rowe, his supervisor, for his encouragement, help, and guidance in the preparation of this thesis and the research which preceded it. He should also like to thank Cathy Penney and Pauline Jones for their valuable and poignant criticisms of the initial draft.

TABLE OF CONTENTS

	Page
LIST OF TABLES	v
LIST OF FIGURES	vi
INTRODUCTION	1
Memorial Representation of Frequency	2
The Spacing Effect	7
The Present Experiment	15
METHOD	17
Materials and Design	17
Subjects	19
Procedure	20
RESULTS	21
Judgments of Frequency	21
Recognition Measures	27
Measures of Accuracy	30
DISCUSSION	35
FURTHER RESEARCH AND THEORETICAL SPECULATION	43
REFERENCES	47
APPENDIX A (Paper Presented at CPA Meeting, 1974)	52
APPENDIX B (Group Instructions)	62

LIST OF TABLES

Page

TABLE 1 Mean Judgments of Frequency by Groups....... 22

TABLE 2 Summary of Analysis of Variance of JOF Results.............................. 25

TABLE 3 Mean Recognition Errors by Groups and Summary of Analyses of Variance............ 29

TABLE 4 Mean Deviations from True Frequencies by Levels of Presentation Frequency......... 31

TABLE 5 Summary of Analysis of Variance of Mean Deviations............................. 33

LIST OF FIGURES

		Page
FIGURE 1	Schematic Representation of the Attachment of List Markers	12
FIGURE 2	Mean Judgments of Frequency for each Group as a Function of Lag and Presented Frequency	23
FIGURE 3	Mean Deviations from Presented Frequencies for each Level of Presented Frequency as a Function of Lag and Group	32
FIGURE 4	Mean Judgments of Frequency for each Level of Frequency as a Function of Spacing and Group from the Preliminary Experiment by Rowe and Rose (1974)	58
FIGURE 5	Mean Judgments of Frequency for each Level of Spacing Collapsed Across Frequency	60

INTRODUCTION

The use of judged frequency as a measure of memory performance with verbal material has become increasingly popular among cognitive psychologists since an important study by Hintzman (1969). In the usual experimental paradigm, subjects are exposed to a fairly large number of words, one at a time for 3-4 seconds each, following which they are asked to judge the frequency of occurrence of each experimental word (along with some "new" words to which the subjects were not exposed). This can be classed as a terminal judgment paradigm. Occasionally a continuous judgment task has been used (e.g. Begg & Rowe, 1972) in which subjects, as they come to each word while studying the list, judge its frequency of occurrence up to that point.

The experiment reported here investigated the effects of three independent variables upon the judgment of frequency in a terminal task. These variables were level of processing, actual frequency of occurrence, and lag (the degree of spacing between repeated words). Briefly, the subjects were required to process the experimental words in one of three incidental tasks, which were assumed to induce different levels of processing. Following this they were unexpectedly asked to judge how often each word in a test list had appeared during the task. The experimental words appeared either 2, 3 or 5 times each during the task and the number of items intervening between repetitions (0, 1, 2, 4, 8, or 16) varied orthogonally with level of presented frequency. Before giving the predictions made prior to the experiment, the introduction will explore current theoretical views touching upon the relevant variables.

Memorial Representation of Frequency

How is frequency represented in memory? Hintzman and Block (1971) presented two hypotheses, one being a trace-strength view which states that the effect of repetition of an event is to increase the strength of a single memory trace representing that event. According to this view, one should only have evidence for the current state of the memory trace and should not be able to make judgments concerning past states or individual presentations of the event. The second hypothesis put forward by Hintzman and Block is the multiple-trace hypothesis which states that each repetition of an event produces a different trace, related to but discriminable from other traces of the same event. They gave evidence from three experiments which showed that subjects can discriminate between different occurrences of the same stimulus to a significant extent. Therefore it appears that, of the two hypotheses, the multiple-trace one is to be preferred.

Anderson and Bower (1972) have postulated a recognition mechanism which incorporates the multiple-trace view. They suggest that, when a word occurs within a study list, a list marker, which summarizes the context in which the word is presented, is attached to the "memory node" corresponding to that word. For each repetition of that word, a new list marker is attached to the node. This model can be used to account for judgments of frequency (JOF). Subjects required to make such judgments can count the number of list markers attached to the memory node in question. Subjects will, of course, make errors in both the attachment and retrieval of list markers but Anderson and Bower nonetheless predict a linear relationship between judged and actual frequency. Underwood, Zimmerman, and Freund (1971) found a relationship between actual and judged frequency which was approximately linear and this has been cited by Anderson and Bower as evidence in support of the application of their model to JOF tasks.

In a recent review of the frequency judgment literature, Howell (1973a) mentions two more possible hypotheses regarding the representation of event frequency in memory. One is the view that both trace strength and number of traces accumulate concurrently, an idea which he calls the multiple-process hypothesis. The strength of a memory trace for a word may be related to the frequency of that word in the subject's linguistic experience, while the number of traces may be related to the frequency of occurrence in the experimental situation. A variation of this hypothesis is that of Underwood (1969a) who suggested that situational frequency is one of several attributes of memory. Underwood does not present a clear view of his concept of trace strength but implies that it is the sum of the values of the stored attributes.

The other possibility put forward by Howell is that the subjects merely count the number of occurrences of each item in the study list. This seems to be done in a continuous judgment paradigm where Begg and Rowe (1972) and Begg (1974) found that the median judgments made by subjects were nearly perfect. One might also expect the use of a counting procedure when the subjects in a terminal judgment task are led to expect a JOF test. However, such does not seem to be the case. Begg (1974) gave some subjects a terminal JOF test following a continuous test while other subjects had only an expected terminal test. If these two groups used similar strategies, one would expect similar results on the terminal tests. The results showed that those subjects who had the initial continuous test judged considerably more closely to the actual frequencies of occurrence than did the subjects given only the terminal test, suggesting that the latter were not using a counting strategy. Also Howell (1973b) found that judgments of frequency were not affected when the subjects were told to expect a recall task as opposed to a JOF task, even for a list containing only ten different words. This independence of set is a further indication that counting is not a prevalent strategy in terminal JOF tests, whether these are expected or not.

Of the four hypotheses mentioned, most evidence favours either the multiple-trace or multiple-process view. However, none of them accounts, without some modification, for a major empirical point made by Begg (1974), namely that a comparison of several studies shows an interesting phenomenon between presented and judged frequency such that those items occurring few times are overestimated while those occurring relatively often are underestimated. Which items are overestimated and which are underestimated depends, not upon their absolute frequency of occurrence in the study list, but upon their relative frequency of occurrence. Begg suggests a sort of "regression to the mean" hypothesis to account for this finding. He states that subjects in a terminal task form a notion of the overall mean frequency of the experimental words. The frequencies of individual words are then judged on the basis of this notion, i.e. the subjects assess the frequency of each word on the basis of its judged deviation from the overall mean. The deviations are judged conservatively, however, such that those words occurring less than the mean number of times are overestimated while those occurring more than the mean number of times are underestimated.

There is some evidence for this view. Begg (1974) found evidence that subjects do form overall mean judgments that are close to the true overall mean. Furthermore, on a terminal JOF test, there was clear overestimation of those words appearing relatively few times and clear underestimation of those words appearing relatively often. In addition, Underwood et al. (1971) found that over a period of a week subjects retain an accurate judgment of the overall mean but the mean judgments of words occurring at different frequencies approach closer to that mean during the period.

However, the "regression to the mean" hypothesis may have to be modified itself or restricted to cases of intentional learning. Rowe and Rose (1974) conducted a terminal JOF test for a list of words presented 2, 3, or 5 times at different levels of lag under three instructional conditions (see Appendix A). An intentionally instructed group were told to remember how often each word appeared, a non-specifically instructed group were merely told to study the words for a forthcoming memory test, while an incidental group were told to rate the words on a dimension of "strength". The data for this experiment are shown in Figure 4. The incidental group, which did not expect the JOF test, gave the highest mean judgments and furthermore overestimated each level of frequency. The other two groups did conform generally to the overestimation-underestimation pattern, although their overall mean judgments differed. The mean for the intentional group was almost exactly equal to the true overall mean frequency while that for the non-specific group was significantly lower than that for the intentional group.

On balance then it appears that the best current model of the memorial representation of event frequency in a terminal paradigm is one of the multiple-trace or multiple-process variety, but with certain modifications required to account for the empirical relationship between actual and judged frequencies of occurrence. The nature of these modifications will remain a mystery, however, until more empirical data is forthcoming. The results of Rowe and Rose (1974) show that the "regression to the mean" hypothesis may be restricted to intentional learning. But even in intentional cases, the evidence is based upon inter-experimental comparisons whereas a real test of the hypothesis requires an intra-experimental comparison.

As for incidental learning, we do not have even cross-experimental comparisons since Rowe and Rose (1974) were the first to use an incidental learning paradigm with a JOF task. What this experiment did show is that judged frequency depends upon pre-test task instructions. Whether for an incidental task it also depends upon the overall mean presented frequency has yet to be tested. If such dependence is not found, then two explanations may be necessary, one for intentional and one for incidental learning. In any event, the existing models will have to be modified.

The Spacing Effect

The spacing effect refers to the increase in memory performance as the level of lag or spacing between repetitions of an item increases. It is an ubiquitous phenomenon, being found not only in JOF tasks (e.g. Hintzman, 1969) but also in studies requiring paired-associate learning (Greeno, 1964), recognition (Hintzman & Block, 1970), and free recall (Melton, 1970). The materials used have included nonsense syllables (Kintsch, 1966), words (Hintzman, 1969b), sentences (D'Agostino & DeRemer, 1973), and pictures (Hintzman & Rogers, 1973).

Hintzman (1974) has presented a lucid review of several hypotheses that have been proposed to explain the spacing effect. The hypotheses can be roughly divided into two categories, those involving voluntary control processes and those involving involuntary processes. An example of the former would be the differential rehearsal hypothesis. This view states that memory performance is better at longer lags because a word repeated with several items intervening between the repetitions receives more rehearsals than one repeated with few or no intervening items. Such differential rehearsal was found by Rundus (1971) who also showed that recall was positively correlated with rehearsal, at least up to a spacing of four items. Differential rehearsal may be sufficient to produce an increase in memory performance over lag but it certainly is not necessary. If it were, then the subjects given intentional learning instructions in the aforementioned study by Rowe and Rose (1974) would have shown a larger effect of spacing than the incidental group. However, as Figure 4 shows, the reverse was the case.

The explanation of the spacing effect currently favoured by most investigators (although not by Hintzman) is the variable encoding hypothesis. Briefly, it states that the same stimulus event can be encoded by a subject in different ways and that the probability of encoding a repeated item in different forms increase as the lag between repetitions increases. Further, the more different forms of encoding that there are, the greater is the probability of retrieving at least one. This explains why recall and recognition increase with increases in lag. However, judgments of frequency also increase over lag. If the variable encoding hypothesis is correct, the variable encoding must also increase the probability of retrieving all list markers, since this is the requirement of a JOF task. Thus this hypothesis not only is an explanation of the lag effect but also enables one to predict that variable encoding imposed upon subjects should lead to increased performance in recall, recognition, and more accurate judgments of frequency.

As Hintzman (1974) points out, one of the difficulties with the variable encoding hypothesis is its ambiguity. It can be (and has been) taken at either of two levels, semantic and non-semantic. When taken at the semantic level, it means that subjects have a tendency to encode the same word with different meanings if there is a relatively long lag between repetitions. While there is some evidence that subjects do encode discretely presented words in a semantic fashion (Light & Carter-Sobell, 1970), there is no direct evidence that repeated words are encoded differentially at relatively long lags. However, some investigators have deliberately manipulated the semantic context of the experimental words by presenting them in phrases during the study trial. If the variable encoding hypothesis is correct at the semantic level, then forcing the subjects to encode the experimental words differentially by presenting repetitions of the words in different semantic contexts should eliminate the lag effect and increase the overall level of performance. This has been done in several experiments involving recall but the results have not been consistent and in some instances have depended upon the presence or absence of cues on the recall trial (e.g. Madigan, 1969). One experiment which did eliminate the lag effect in a differential encoding paradigm and increased the probability of recall dramatically was that of Gartman and Johnson (1972). However, this study is unusual, not only for the large effect found with differential encoding, but also for its procedure which employed relatively few experimental items whose semantic contexts were set by the items preceding them in the list and not by phrases. In a JOF paradigm, Rowe (1973a,b) employed homonyms in repeated phrases and different-meaning phrases. Since lag was randomised in these studies, they shed no light on the spacing effect but Rowe did find a <u>decrease</u> in mean JOF with different-meaning items relative to same-meaning ones, which is contrary to the prediction based upon the differential encoding hypothesis.

Taken on a non-semantic level, variable encoding can be depicted by the model of Anderson and Bower (1972). They conceive of a list marker as a bundle of contextual cues which accompany the presentation of a word (including one which states that the word is a member of the set of words defining the list). When the repetitions of a word are successive or nearly so, then the contextual cues in the succeeding list markers are very similar such that the list markers "overlap". This is schematized in Figure 1(a) and in such situations it would be difficult to discriminate all list markers. Hence the judgments of frequency at short lags would be low. With longer lags between repetitions, the contextual cues accompanying each presentation would have less overlap, thus producing more discriminable list markers. This situation is schematized in Figure 1(b). As pointed out by Hintzman (1974), one of the problems with the variable encoding hypothesis at the non-semantic level is that the characteristics of the cues have yet to be established. They do not include differences in modality since such differences do not attenuate the lag effect (Hintzman, Block, & Summers, 1973). The results of Gartman and Johnson (1973) might lead one to suspect that list context is a contributing factor. Hintzman (1974) reports a hitherto unpublished study in which the experimental words were repeated in the same list context in one condition or in different contexts in another condition and did find a slight but significant effect of context upon JOF. However, judgments were higher in the same context condition and this is contrary to the differential encoding hypothesis (c.f. Rowe, 1973a,b).

The variable encoding hypothesis, at least at the semantic level, is another member of the class of explanations of the lag effect involving voluntary processes. An example of an explanation involving involuntary processes is the consolidation hypothesis which states that, if a word

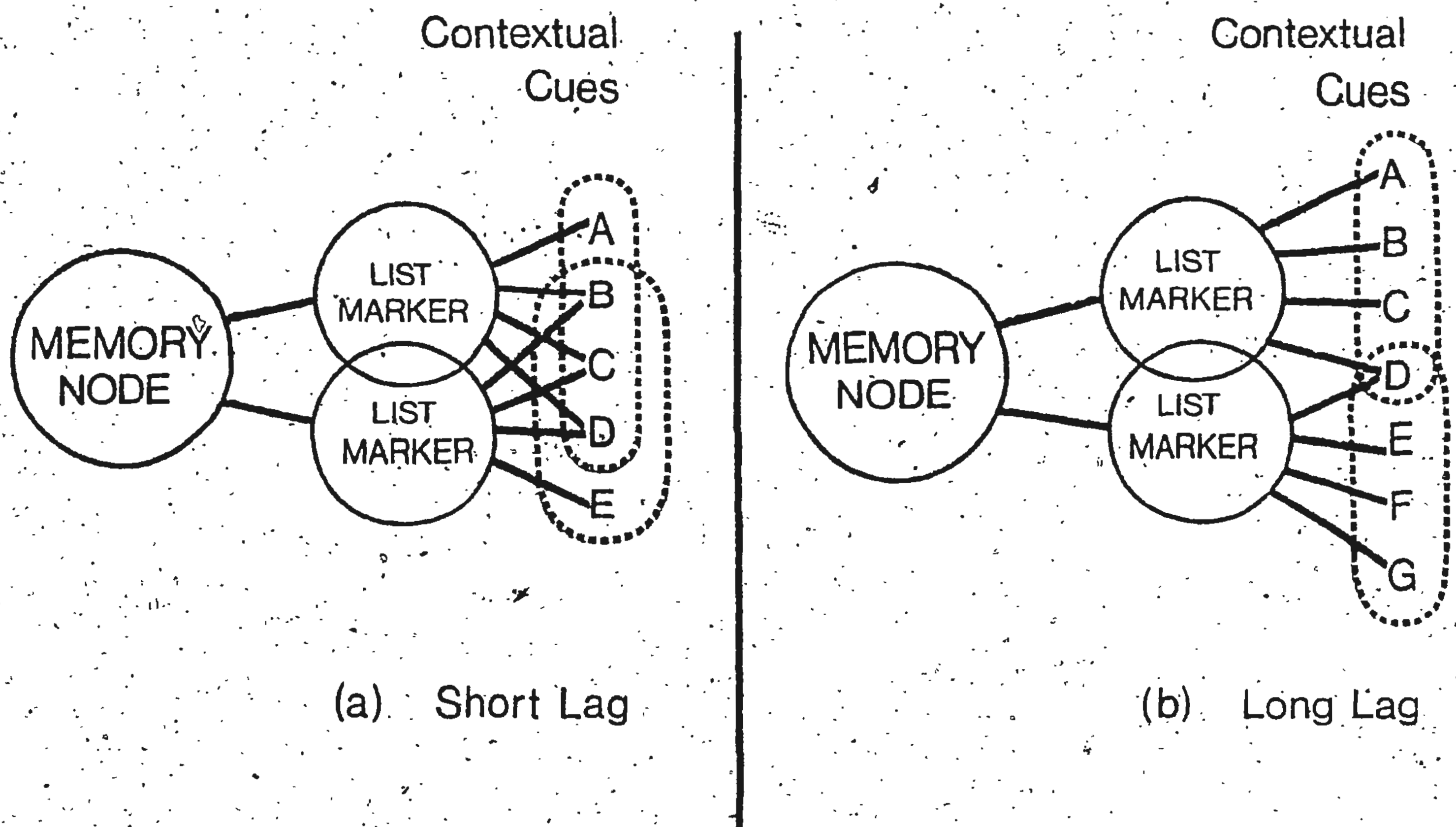


Figure 1: Schematic representation of the attachment of list markers

is repeated before the memorial representation of its first occurrence is consolidated, then the memory of the first occurrence is weak. Thus memory performance for events repeated at short lags is low because encoding or storage of initial presentations is poor. However, Hintzman et al. (1973) give evidence that the locus of the lag effect lies not with the initial presentations but with retrieval of later ones.

Although Hintzman and his co-workers found evidence against the consolidation hypothesis, he does favour an explanation involving involuntary processes. His view, which he calls an habituation hypothesis, is that when a stimulus event occurs, its memorial representation becomes habituated or adapted, i.e. its threshold for re-activation is raised. Recovery from this habituation begins as soon as the stimulus ceases but it takes time. If a second presentation of the stimulus event occurs before recovery is complete, then encoding and subsequent retention of that second presentation will be poor. Hintzman expands this idea to state two implications concerning the role of rehearsal. First, rehearsal maintains habituation and hence enhances the lag effect. Hintzman concluded from a survey of several studies that the lag effect generally occurred when a repetition came within 15 seconds of the preceding presentation of a word. However, most of these studies (including his own) used non-specific instructions. Where specific instructions are used, one would expect a spacing effect over a longer time interval because there should be more rehearsal. The second "corollary" states that, when rehearsal does not occur, recovery should depend only upon time.

The experiment by Rowe and Rose (1974) has important bearings upon the habituation hypothesis. A glance at Figure 4 (Appendix A) shows that the non-specific group had a tendency to produce peak judgments at a lag of four, which in that experiment occurred at a time interval of 16 seconds and corresponds well to Hintzman's limit of 15 seconds for the spacing effect. The specific group, who might be assumed to rehearse more than the non-specific one, did show a prolongation of the lag effect, at least for items presented five times. The fact that the incidental group showed a spacing effect also supports the general view that this effect is due to involuntary processes. However, the finding that the incidental group showed the greatest and most prolonged effect of spacing goes against the two implications of Hintzman's habituation hypothesis because, according to it, the most prolonged lag effect should have been shown by the group whose instructions would induce the greatest amount of rehearsal. Instead it was shown by the group whose instructions should have induced little or no rehearsal. Secondly, the results for the incidental group show that, when rehearsal does not occur, the lag effect depends not only upon time but also upon the level of presentation frequency.

To summarize, the spacing effect is explained by most investigators as due to differential encoding at the semantic level. However, there are at present empirical findings that are inconsistent with this hypothesis. Hintzman (1974) argues that hypotheses involving involuntary processes need consideration and he proposes that the spacing effect is due to adaptation or habituation of the memorial representation of a word. The experiment by Rowe and Rose (1974) supports in principle the view that the spacing effect is due to involuntary processes but it does not support specifically the habituation hypothesis.

The Present Experiment

The experiment reported here extended the study by Rowe and Rose (1974) by examining the effects of frequency and lag upon judgments of frequency under three types of incidental instructions, designed to induce different levels of processing of the presented words. The concept of "level of processing" stems from Craik and Lockhart (1972). They postulate a hierarchy of processing stages involved in the encoding of verbal material, beginning with awareness of stimulation through levels of perceptual analysis to what has traditionally been called short-term memory and long-term memory. The implication is that the depth of a level of processing varies directly with the degree of cognitive or semantic analysis. Furthermore, memory trace persistence is assumed to be a positive function of depth of processing.

Level of processing was included as a variable in this experiment in order to examine its effect upon JOF in interaction with presented frequency and lag. One purpose of the experiment is to corroborate the finding by Rowe and Rose (1974) that certain incidental tasks lead to a prolongation of the lag effect and to high levels of memory performance. As mentioned earlier, the incidental group in that experiment had to rate each word for degree of "strength". To carry out such a task, the subjects must consider the connotative meanings of each word and hence semantic processing is involved. If the levels of processing view is correct, then it seems that the incidental group processed the words to a deeper semantic level than the two intentional groups, since they produced the greater judgments. Furthermore, it seems that the lag effect is enhanced, not necessarily by rehearsal, but by deeper processing.

The present experiment attempts a direct test of this hypothesis by using three incidental tasks. This was considered to be an important modification of the earlier experiment from the conceptual viewpoint of level of processing because there is more experimental control with incidental learning than with intentional learning. One group (the evaluation group) was asked to evaluate each word in the list for its implied "goodness", a second group (the rhyming group) was asked to find a rhyme for each experimental word, while the remaining group (the coding group) had to code each word arithmetically according to values arbitrarily assigned to vowels by the experimenter. Since the evaluation of "goodness" and the evaluation of "strength" are similar semantic tasks, it was predicted that the JOF results for the evaluation group here would be very similar to those for the incidental group of Rowe and Rose (1974).

If frequency judgments are affected by the level of processing, the coding group should produce lower judgments than the evaluation group. Rowe (1974) has recently suggested that frequency judgments decrease with a non-semantic as compared to a semantic level of processing and the same result was expected here. The rhyming group, whose level of processing is assumed to be intermediate between the other two, should produce frequency judgments of an intermediate value.

As already stated, the judgments for the evaluation group should conform closely to those for the incidental group of Rowe and Rose (1974). They should then show a large effect of spacing. If the assumption concerning the depth of processing of the rhyming task is correct and if the lag effect is in fact enhanced by deeper processing, then we should expect the evaluation group to exhibit the greatest effect of spacing with the coding group showing the least effect and the rhyming group in between.

METHOD

Materials and Design

The items used in this experiment were common words (frequency A or AA from the word count of Thorndike and Lorge, 1944) of one or two syllables and five or six letters each. The experimental words appeared in the presentation list with a frequency (F) of two, three, or five. The lag (L) between repetitions was varied orthogonally with F and had six levels: 0, 1, 2, 4, 8, and 16. This yielded a factorial design of 18 FxL cells. Five different words were allotted to each cell with no word appearing in more than one cell.

The words were typed singly in lower case on three x five inch white cards. The pack of cards was divided into quintiles and in general one word from each cell, together with its repetitions, was allocated to each quintile. (This procedure broke down only for words repeated five times with a lag of 16, which have a range of 65 items, whereas a quintile contained approximately 64 positions. Only one of the five occurrences of a word in this cell occurred outside the allotted quintile.) Besides the 90 experimental words, there were 27 additional words which appeared once only. These items served as "fillers" to create the proper degree of spacing between repetitions and also occupied the first four and last four positions in the pack as a control for primacy and recency effects.

The words were assigned a position within the first quintile quasi-randomly. The numbers 1 to 18 were allocated randomly as labels to the 18 FxL cells and then these numbers were again randomized to determine the order in which the cell conditions would appear in the pack. However, the resulting order was only quasi-random because an effort was made to fit those items having short lags between the repetitions of those items having long lags. Once the first words allocated to each FxL cell had been assigned positions in the pack, the randomized order was gone through four more times until all five words in each cell had been assigned. In spite of the modification of the randomizing procedure, it was not possible to fill every space between the repetitions with experimental words and so some "blank" positions remained. Nineteen single words were assigned to these positions.

In order to counterbalance any effects on lag attributable to individual words, every word appeared once at each of the six levels of L. This necessitated the construction of six different packs of cards. Once the first pack had been constructed, the remaining packs were made from it by keeping the positions of the fillers constant while rotating the sets of five words from one FxL cell to another. Since there were three levels of F, any interaction between presentation frequency and individual words was controlled by rotating in such a manner that each word appeared twice at each level of F. For instance, the five words appearing in pack 1 at F=2, L=0, appeared in packs 2 to 6 at the following levels respectively: F=3, L=1; F=5, L=2; F=2, L=4; F=3, L=8; F=5, L=16. Thus the positions occupied by each FxL cell (and by the fillers) were fixed once the first pack was constructed. What varied from pack to pack was the set of five words occupying each cell.

Subjects (Ss)

In total, 146 Ss, all of whom were undergraduates at Memorial University and paid $2.00 for their participation in the experiment, were tested. Two of these were discarded for failing to obey instructions, so the data for 144 Ss are reported.

Procedure

Equal numbers of Ss were allocated to each of three groups: the evaluation group, the coding group, or the rhyming group. As was mentioned previously, the Ss in the first group were told to rate each word for its connotative "goodness" on a seven-point scale, while those in the rhyming group had to find a rhyme for each word. The Ss in the coding group had to add up for each word the arbitrarily assigned arithmetical values of the vowels. These values were: a=1=u, e=2=o, i=3. The Ss were told that some words would be repeated in the pack and they were given examples of their incidental tasks, but they were not told about the terminal JOF test. The full instructions given to each group are shown in Appendix B.

The Ss were tested in sub-groups of four to six with successive sub-groups being assigned to tasks in such a manner that the number tested under one condition at no time exceeded the number tested under any other condition by more than six. Eight Ss in each of the three groups were assigned to each pack of cards. After they had read the instructions provided on a separate sheet for each S, they were allowed to ask questions and were given a few practice items to process according to their incidental task in time to an electronic metronome which sounded once every four seconds. If there were no further questions, the Ss went through the pack of cards in time to the metronome. The Ss recorded their responses in writing on a response sheet. In order to counterbalance any interaction of order of occurrence within a pack with individual words, half of the Ss in each group went through their packs in reverse order.

After the Ss had gone through the pack of cards, they were given a list of 90 experimental words with 30 new words placed randomly among them. The 120 test words appeared in 24 sets of five, formed into four columns of six sets each on a single page. The Ss were asked to judge how often each word appeared in their pack of cards. The test was unpaced. In a previous, unpublished study (Rose, 1973), it was found that the number of false alarms, i.e. the number of new words given non-zero judgments in a JOF test, increased as Ss went through the test. In order to control this type of response bias, half of the Ss went through the test in reverse order. The order of answering the test varied orthogonally with the direction of going through the pack of cards. The entire experiment took approximately 40 minutes.

RESULTS

Judgments of Frequency

The means for each group (G)xFxL cell are given in Table 1 and are depicted graphically in Figure 2. A three-way analysis of variance was performed on the data, the

TABLE 1

Mean Judgments of Frequency by Groups

FREQUENCY	LAG 0	1	2	4	8	16	$\bar{X}$
				CODING			
2	2.42	2.54	2.69	2.73	2.52	2.55	2.57
3	3.31	3.20	3.00	3.09	2.88	3.20	3.11
5	3.28	3.71	3.44	3.52	3.49	3.64	3.51
$\bar{X}$	3.01	3.15	3.04	3.12	2.96	3.13	3.07
				RHYMING			
2	2.29	2.83	2.89	3.13	3.20	3.26	2.93
3	3.29	3.48	3.77	4.21	4.20	4.64	3.93
5	3.66	4.66	4.73	4.96	5.34	6.05	4.91
$\bar{X}$	3.09	3.66	3.80	4.10	4.25	4.65	3.93
				EVALUATION			
2	2.66	3.37	3.68	3.43	3.59	3.68	3.40
3	3.28	3.58	4.07	4.27	3.97	4.43	3.93
5	3.84	4.79	4.72	5.18	5.71	6.12	5.06
$\bar{X}$	3.26	3.91	4.16	4.29	4.42	4.74	4.13
$\bar{X}$	3.12	3.57	3.66	3.84	3.88	4.18	3.71

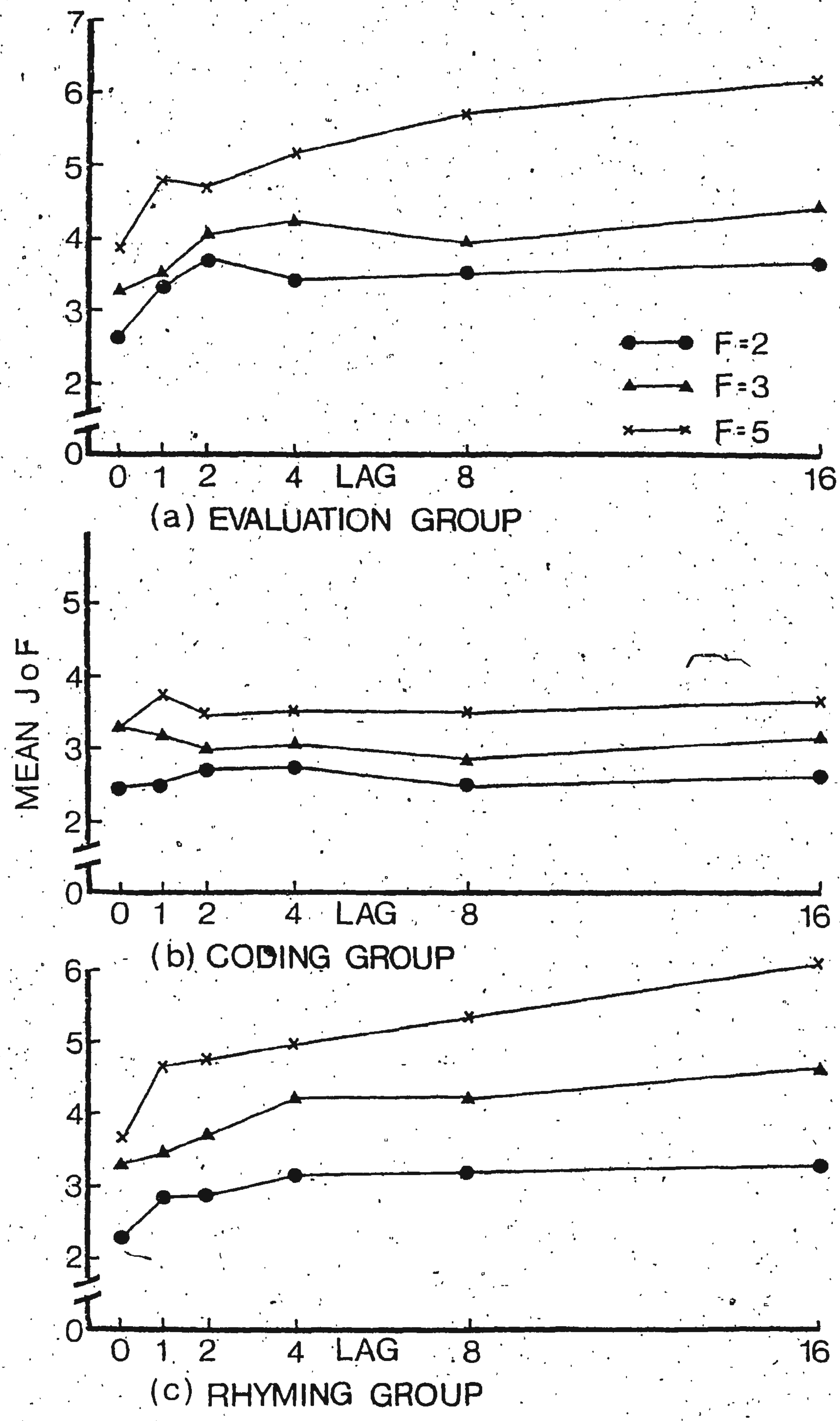


Figure 2: Mean judgments of frequency for each group as a function of lag and presented frequency

summary of which appears in Table 2. There was a main effect of groups, $F(2,141) = 6.41$ ($p<.01$). The other two factors, frequency and lag, were also significant ($p<.001$) with $F(2,282) = 253$ and $F(5,705) = 36.0$ respectively. A comparison of individual group means by the Newman-Keuls test showed that the evaluation and rhyming groups did not differ significantly but both of these showed significantly higher judgments than the coding group ($p<.01$). All three levels of frequency differed significantly among themselves ($p<.01$).

All of the double interactions were significant ($p<.001$). For the GxF interaction, $F(4,282) = 12.1$; for GxL, $F(10,705) = 8.34$; for FxL, $F(10,1410) = 4.94$. Only the triple interaction was not significant. The GxF interaction reflects the tendency for the coding group to show relatively small increases in JOF as actual frequency of occurrence increases. The GxL interaction similarly appears to reflect the relatively small effect of lag on JOF for the coding group while the FxL interaction reflects an overall tendency for the lag effect to increase with higher levels of F.

The second of these tendencies was checked by analyzing the results for each instruction group with separate two-way analyses of variance. For the coding group, only frequency showed a significant effect, $F(2,94) = 36.5$ ($p<.001$). The F-ratio for the variable of lag was less than one. For the rhyming group, both frequency, $F(2,94) = 110$, and lag, $F(5,235) = 19.95$, were significant ($p<.001$) but the interaction was not. A Newman-Keuls test was applied to the

TABLE 2

Summary of Analysis of Variance of JOF Results

Source	df	MS	F
Between Subjects			
Instructional group (G)	2	6875.96	6.41*
Ss within groups (S)	141	1073.43	
Within Subjects			
Presentation frequency (F)	2	12614.1	253.2**
G x F	4	603.03	12.11**
F x S	282	49.82	
Lag (L)	5	1361.08	36.04**
G x L	10	314.85	8.34**
L x S	705	37.77	
F x L	10	163.58	4.94**
G x F x L	20	34.14	1.03
F x L x S	1410	33.12	

*$p < .01$
**$p < .001$

individual means for lag collapsed across frequency. This was done because some investigators have found no significant differences among non-zero levels of lag (e.g. Underwood, 1969b) and have merely distinguished between "massed presentations" (lag = zero) and "distributed presentations" (lag greater than zero). The test showed that the JOF at lag = 0 was significantly below all others ($p<.01$) while at lag = 16 it was significantly above all others ($p<.05$).

For the evaluation group all factors were significant and, since the FxL interaction was significant, $F(10,470) = 6.40$ ($p<.001$), one-way analyses were carried out for each level of frequency. Each level showed a significant effect of lag with $p<.001$. At F = 2, $F(5,235) = 9.37$; at F = 3, $F(5,235) = 11.7$; at F = 5, $F(5,235) = 28.4$. Newman-Keuls comparisons of individual means showed that at F = 2, the mean judgment at lag = 0 was significantly below all others ($p<.01$). At F = 3, the mean JOF's at lag = 0 and lag = 1 were significantly below all others ($p<.01$), while at F = 5, the mean JOF at lag = 0 was significantly below all others ($p<.01$) and those at lag = 8 and lag = 16 were significantly above all others ($p<.01$).

In summary then, we find that the type of incidental task does affect the performance on a terminal JOF test, although the mean frequency judgments for the rhyming group were higher than expected. The type of incidental task also interacted with frequency and with lag. The coding group showed no significant lag effect while, in terms of range of

JOF over the six levels of spacing, the evaluation group showed the greatest effect. It was also found that the lag effect interacted with the level of actual frequency, being greatest at F = 5, although this was significant only in the case of the evaluation group. Finding significant effects of lag among non-zero levels of lag for the rhyming group in general and for the evaluation group at F = 5 was important in corroborating the concept of spacing effect per se, as opposed to a mere distinction between massed and distributed presentations. Finally, the results for the evaluation group were very similar to those for the incidental group of Rowe and Rose (1974), as predicted. In fact, the rank correlation between the cell means of the two groups was +.98 ($p<.001$).

Recognition Measures

Measures of recognition of the "yes-no" variety can easily be derived from JOF results. Any experimental word given a non-zero judgment has obviously been correctly recognized as having occurred in the pack and can be designated a hit. Conversely, any experimental word judged as occurring zero times constitutes a miss while a new word on the test given a non-zero judgment is a false alarm.

It is intuitively reasonable to expect measures of recognition and judgments of frequency to be positively related and indeed there is evidence for this (e.g. Hintzman & Block, 1970). However, such does not have to be the case. It is possible for a group of Ss to "miss" a large number of words and at the same time exaggerate the judged frequency of their hits such that the mean JOF is reasonably accurate. To check this possibility, measures of recognition performance and accuracy were taken and analysed.

One-way analyses of variance were carried out for misses, false alarms, and combined errors (i.e. misses + false alarms) with G being the main factor. The reason for collapsing across lag and frequency levels to do only a one-way analysis is because false alarms (and hence combined errors) cannot be ascribed to FxL cells as can misses. The mean recognition errors and the F-ratios for each analysis are shown in Table 3. The three groups were mutually separable only for combined errors, $F(2,94) = 101$ ($p<.001$). An analysis of individual means by the Newman-Keuls test showed that the evaluation group made significantly fewer errors than the rhyming group which in turn made significantly fewer errors than the coding group ($p<.01$). In the analysis of misses, $F(2,94) = 31.05$ ($p<.001$) with the evaluation group making fewer errors than the other two groups ($p<.01$) which did not differ from each other. In the analysis of false alarms, $F(2,94) = 79.8$ ($p<.001$) with the coding group making significantly more false alarms than the other two groups ($p<.01$) which did not differ significantly from each other.

The number of hits for each GxFxL cell were tallied but were not analyzed, because the evaluation group had such a high performance (nearly perfect in some cells) that they did not generally discriminate among the three levels of frequency and so produced a "ceiling" effect. A test of the

TABLE 3

Mean Recognition Errors by Groups and Summary of Analyses of Variance

(a) Mean combined recognition errors (misses + false alarms)

Group	Mean	Analysis of Variance Summary			
		Source	df	MS	F
Evaluation	8.58				
Coding	36.21	Group (G)	2	9162	101*
Rhyming	21.88	Subjects (S)	94	90.5	

(b) Mean recognition misses

Group	Mean	Analysis of Variance Summary			
Evaluation	3.54	Source	df	MS	F
Coding	19.85	G	2	3534	31.1*
Rhyming	16.31	S	94	114	

(c) Mean recognition false alarms

Group	Mean	Analysis of Variance Summary			
Evaluation	5.04	Source	df	MS	F
Coding	16.35	G	2	1958	79.8*
Rhyming	5.56	S	94	24.5	

*p<.001

homogeneity of variance among the 54 cells showed that F_{max} exceeded 70 which is far above the critical value of approximately 3.5 at the 1% level of significance. In general, the coding and rhyming groups showed increased performance with an increase in presentation frequency and the rhyming group showed an effect of lag. The similarity between recognition and JOF performances was substantiated by a correlation of +.49 between the number of hits per cell and the mean judgment of frequency for that cell.

In summary then, we find that the recognition performance of the groups is close to that found in the JOF data. One cannot attribute the results of the evaluation group to a recognition response bias since they showed the fewest errors as well as the most hits. Since the rhyming group made significantly more errors than the evaluation group but were not significantly below them in mean judgments, it seems that they exaggerated their judgments of those words correctly recognized. This possibility was investigated in the next section.

Measures of Accuracy

The accuracy of each judgment made by each S was taken to be its absolute difference from the actual frequency of occurrence for that word. The mean deviations are given in Table 4 and are depicted graphically for each separate level of frequency in Figure 3. A three-way analysis of variance was carried out on these and is summarized in Table 5. Group was a significant factor with $F(2,141) = 4.27$

TABLE 4

Mean Deviations from True Frequencies by Levels of Presentation Frequency

	LAG						
GROUP	0	1	2	4	8	16	$\overline{X}$
				F = 2			
Evaluation	1.43	1.76	1.95	1.84	1.84	1.99	1.80
Coding	1.88	1.80	1.99	1.98	1.76	1.78	1.86
Rhyming	1.90	1.96	2.08	2.25	2.19	2.25	2.11
$\overline{X}$	1.74	1.84	2.05	2.02	1.93	2.00	1.93
				F = 3			
Evaluation	1.64	1.62	1.73	1.69	1.78	1.96	1.74
Coding	2.23	2.23	2.12	2.20	2.03	2.15	2.16
Rhyming	2.42	2.08	2.27	2.57	2.46	2.71	2.41
$\overline{X}$	2.10	1.98	2.03	2.15	2.10	2.28	2.11
				F = 5			
Evaluation	2.42	1.86	2.10	2.03	2.10	2.16	2.11
Coding	2.92	2.82	3.12	2.92	2.87	2.90	2.93
Rhyming	2.82	2.86	2.81	2.56	2.58	2.72	2.72
$\overline{X}$	2.72	2.52	2.68	2.51	2.51	2.59	2.59
$\overline{X}$	2.18	2.11	2.25	2.23	2.18	2.29	2.21

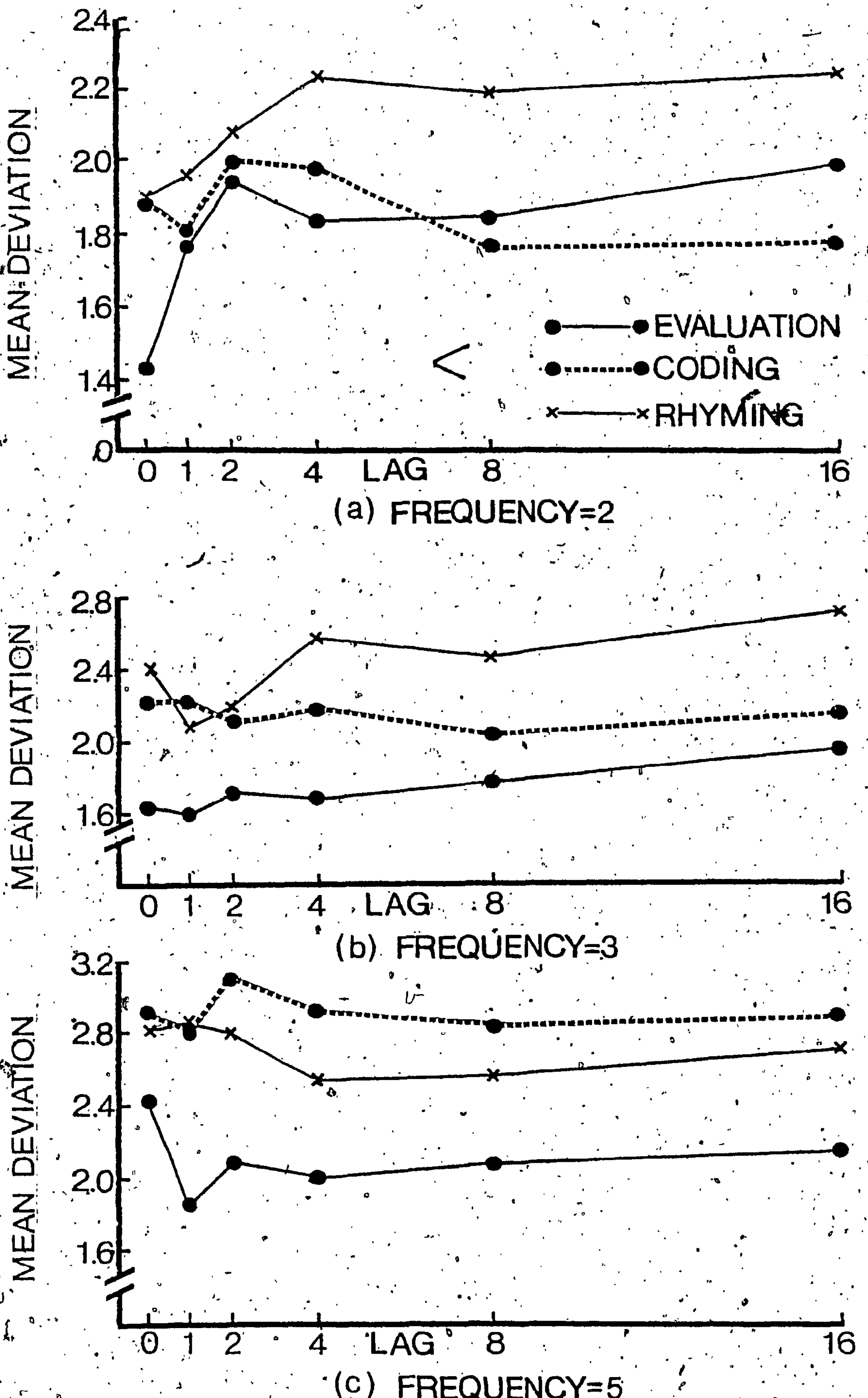


Figure 3: Mean deviations from presented frequencies for each level of presented frequency as a function of lag and group

TABLE 5

Summary of Analysis of Variance of Mean Deviations

Source	df	MS	F
Between Subjects			
Instructional group (G)	2	1764.6	4.27*
Ss within groups (S)	141	413.3	
Within Subjects			
Presentation Frequency (F)	2	2515.3	61.68**
G x F	4	293.7	7.20**
F x S	282	40.78	
Lag (L)	5	44.83	1.99
G x L	10	20.20	0.90
L x S	705	22.50	
F x L	10	40.78	2.05*
G x F x L	20	22.69	1.14
F x L x S	1410	19.93	

*$p < .025$
**$p < .001$

($p<.025$). Individual comparisons of the means showed that there were no differences between the rhyming and coding groups while the evaluation group had significantly lower deviations than the other two ($p<.01$). Frequency was also a significant factor, $F(2,282) = 61.7$ ($p<.001$). Individual comparisons of the means showed that the mean deviation increased directly with the level of frequency with the three means being significantly different from each other ($p<.01$).

Lag was not a significant factor nor were the GxL nor GxFxL interactions. However, the GxF interaction was significant, $F(4,282) = 7.20$ ($p<.001$). A glance at Figure 3 indicates that the root of this is the switch in the rank order of the means for the coding and rhyming groups as F increases from three to five. Two-way analyses of the deviations were carried out for each level of frequency separately. At $F = 2$, there were no differences among the three groups, $F(2,141) = 1.09$, but at $F = 3$ there was a significant difference, $F(2,141) = 4.94$ ($p<.01$). A Newman-Keuls comparison of the individual means showed that the deviation for the evaluation group was significantly lower than that for the rhyming group ($p<.01$). At $F = 5$, G was a significant factor, $F(2,141) = 8.50$ ($p<.001$) with the evaluation group being significantly lower than the other two groups ($p<.01$) which did not differ significantly from each other. Only at $F = 2$ was there a significant lag effect, $F(5,705) = 3.26$ ($p<.01$) with the mean deviation at lag = 0 being below the means at lags 2, 4, and 16 ($p<.05$). There were no significant GxL interactions and so no one-way analyses were carried out.

Finally, there was a significant FxL interaction in the three-way analysis, $F(10,1410) = 2.05$ ($p<.03$). This reflects the change in the mean deviation at lag = 0, which is the lowest at F = 2 but increases as F increases to be the highest lag mean at F = 5.

In summary we see that the evaluation group produced not only the highest mean judgments of frequency and the highest mean number of correct recognitions but also the most accurate judgments. The coding and rhyming groups were not separable on deviations. However, it is felt that this is due to the narrow range of levels of frequency used. The coding group showed a tendency to be conservative in their judgments compared to the other two groups. This is advantageous at low levels of frequency where the general tendency (in other experiments as well as this one) is to overestimate. However, this is disadvantageous at high levels of frequency where the general tendency is to underestimate anyway. It seems reasonable to suggest that the relative positions of the coding and rhyming groups for mean deviations at F = 5 would be even more exaggerated at higher levels of frequency.

DISCUSSION

The experiment reported here leads to two important general conclusions, one of which is that the judged frequency of occurrence of a word increases as the word is processed to deeper levels. This follows from the finding that, as predicted, the coding group produced significantly lower judgments than the evaluation group. The JOF measures by themselves could, of course, reflect no more than some tendency to overestimate inaccurately. However, as we have seen, the evaluation group not only made the highest judgments but were most accurate. Hence we are justified in concluding that the group which processed the words to the deepest level showed the best memory performance.

In this the results are in line with the work of Jenkins and his colleagues (Jenkins, 1974) using free recall of highly associated words. They found that incidental tasks requiring semantic processing produced greater recall than those requiring non-semantic processing. As Rowe and Rose (1974) found for their incidental group, they discovered that semantic orienting tasks serve as well as or better than strategies used by subjects who intend to learn. Thus the experiment reported here and its preliminary one support the conclusions of Jenkins that memory "is not an automatic consequence of having intended to remember.... (nor) of having responded to events" but is a consequence of the kind of response that has been made.

However, the results of the rhyming group here do depart from those of Jenkins and his co-workers. They found that a rhyming task is no more effective in prompting recall than are other non-semantic tasks, whereas here the mean JOF of the rhyming group was not significantly below that of the evaluation group. There are several possible reasons for this discrepancy. One is that the judgments of the rhyming group relative to those of the evaluation group are the result of a combination of fewer correct recognitions on the one hand and exaggerated judgments of the hits on the other hand. The similarity of the mean judgments of the two groups may have been somewhat fortuitous therefore, and one may conclude that the rhyming task is in fact less effective than the semantic task. A second point to be made is that the comparison made by Jenkins between rhyming and other non-semantic tasks is a cross-experimental one. Johnston and Jenkins (1971) compared subjects given a rhyming task with subjects given a semantic task and the latter showed clear superiority. However, no experiment seems to have included semantic, rhyming, and other non-semantic tasks together. If this were done, it might be found that rhyming produced results intermediate between the semantic and other non-semantic tasks.

As a third reason for the discrepancy, it could be argued that the groups in this experiment were differentially efficient in carrying out their incidental tasks and this factor could have been confounded with the level of processing. Indeed, it was found that while the evaluation group omitted on average less than one item from the 327 words in a pack, the coding group omitted 5.5 and the rhyming group omitted 82.4. These large differences are a reflection of the choice of words used, which came randomly from the

Thorndike-Lorge word count. Thus they were not chosen for ease of rhyming. If we divide the 48 Ss in the rhyming group into two sub-groups according to whether their omissions in the incidental task numbered above or below the median (which was 87), we find that those Ss who had a higher number of omissions also produced higher mean judgments of frequency. The mean JOF's for the two sub-groups were 4.45 and 3.40 respectively which differed significantly according to the Mann-Whitney test ($Z=2.64$, $p<.005$). Hence the mean JOF of the rhyming group could have been inflated because of relative failure to carry out their incidental task and this could be cited as a methodological weakness of the present experiment. Another set of words for which rhymes could be found easily might then induce judgments of frequency in a rhyming group which were significantly below those for an evaluation group. However, one can counterargue that the main factor is not success in the task but processing the words to an appropriate level. A subject asked to find a rhyme for a word within four seconds should be attentive to the phonetic qualities of that word, whether he succeeds in the task or not. On balance, though, it must be admitted that no firm conclusion can be made concerning the level of processing induced by the rhyming task in this experiment.

As predicted in the introduction, the judgments of frequency for the evaluation group were very similar to those for the incidental group in the preliminary experiment by Rowe and Rose (1974). This then substantiates the argument that Begg's "regression to the mean" view may have to be restricted to intentional tasks. The hypothesis cannot be ruled out for incidental tasks, however. It is possible that Ss in such conditions form their judgments from their overall mean JOF which is considerably higher than true mean F. Some regression toward this mean could still produce overestimation of all levels of F.

The second general conclusion from this experiment is that the lag effect is enhanced by deeper processing. This is seen in the evaluation group who processed the words most deeply and showed the most prolonged effect of spacing, especially at a frequency of five. The interesting finding here is that the coding group showed no effect of lag at all. This is evidence that a lag effect will occur only when there is processing beyond the superficial level used by the coding group. If the coding group carried out their task literally, they did not need even to look at the words and in this case it would not be surprising to find no consistent spacing effect. However, the coding group did show a significant increase in judged frequency as presentation increased and this indicates that they did notice the words to an extent sufficient to affect the JOF. The depth of processing required for an effect of spacing seems to be greater than that required to make judgments of frequency.

One explanation of this finding could be that the evaluation of rhyming groups deal with the words wholistically while the coding group processes merely parts of the words, namely the vowels. This suggestion was made to Jenkins after the early experiments by him and his colleagues showed that semantic incidental tasks led to better recall than orthographic tasks (see Jenkins, 1974). His refutation of this argument was to show that a wholistic rhyming task also induced a low-level of recall performance. Our results indicate that a rhyming task may not be the low-level process which Johnston and Jenkins (1971) found it to be and so the "part-whole" suggestion may have some validity. However, it is difficult to see how subjects who only process parts of words can discriminate the frequency of occurrence of those words to a significant extent.

A better explanation may be one which refers again to the discriminability of list markers. It is probable that the subjects in the coding group glanced at most, if not all, of their words but that the wholistic processing was so superficial that the list markers attached to the corresponding memory nodes were weakly discriminable. This could occur because, with less processing of the words wholistically by the coding group, fewer contextual cues would be included in each list marker (in comparison to the other two groups). The fewer the contextual cues, the less is the probability that one list marker can be distinguished from another. In other words a situation like that depicted in Figure 1(a) would hold for any lag between repetitions and spacing would have no significant effect. If this explanation holds, then there must be some other dimension such as strength by which judgments of frequency are made and to which the coding group add a significant (albeit relatively small) increment each time they process a word.

Aside from the two general conclusions, the experiment reported here has implications for existing hypotheses. Mention has already been made of the restrictions it places upon Begg's "regression to the mean" hypothesis concerning the memorial representation of event frequency. The corroboration of the prediction that the results for the evaluation group here would be similar to those for the incidental group of Rowe and Rose (1974) adds weight to the argument against Hintzman's habituation hypothesis. While the finding of a pronounced lag effect with incidental learning fits in with the general view that voluntary control processes are not involved, it does not support the two corollaries to Hintzman's hypothesis. The two experiments show that the spacing effect in non-rehearsal situations is not limited by time and the effect is not prolonged in a rehearsal situation relative to a non-rehearsal one. The experiment also provides clear evidence of a spacing effect extending beyond a lag of zero. Some investigators have argued for a distinction between the "lag effect" and the "massed-versus-distributed practice effect" (c.f. Hintzman, 1974, pages 79-80) but the results here vindicate those who believe that it is more parsimonious to combine the two phenomena.

Finally there are empirical implications made by this experiment and the preliminary one upon which it is based which substantiate the view of Howell (1973a) who states that "(JOF) experiments to date may have focussed too narrowly" (page 46). Considerable mention has already been made of the importance of looking at more than the main dependent variable in JOF studies. Secondly, these experiments have shown the theoretical importance of considering the type of instructions given to subjects. As mentioned in the introduction, Hintzman (1974) concluded that the lag effect occurred over an interval of about 15 seconds. Only one of the six groups in the two experiments showed any evidence of this and that was the group given non-specific instructions, precisely the type of instructions used by Hintzman and his colleagues in their experiments. Another important point corroborated by these experiments is the interaction between lag and frequency. This interaction has been found before (e.g. Underwood, 1969b) and Hintzman (1974) takes note of it in his review (page 79). However, it has been ignored theoretically, probably because many experiments involving spacing as a variable have used frequencies of only one and two (e.g. D'Agostino & DeRemer, 1973; Gartman & Johnson, 1972; Rundus, 1971; Melton, 1970). Any adequate explanation of the spacing effect must account for the enhancement of the effect as actual frequency increases.

In summary then, this experiment has shown that, judgments of frequency depend upon the level of processing, with the deepest level producing the highest and most accurate judgments. Even if one is conservative and ignores the level of processing involved with the rhyming group, there is clear separation between the evaluation group and the coding group which supports this. The results also show that the level of processing interacts with the effects of spacing, with the deepest level producing the most pronounced lag effect. These findings have implications for existing hypotheses concerning the spacing effect and the representation of event frequency in memory as well as for the methodology employed in studies of these phenomena.

FURTHER RESEARCH AND THEORETICAL SPECULATION

There are at least two experiments suggested by this study and its theoretical orientation. Mention was made in the Introduction that Begg's (1974) idea, which has been called here a "regression to the mean" hypothesis, is based upon inter-experimental comparisons. What is really required is an intra-experimental comparison with one group of subjects being presented with words, say, 2, 3, and 5 times and a second group being presented with the same words at frequencies of say, 2, 5, and 8. The main interest would center on the words presented five times to each group, because these should be underestimated by the first group relative to the second, if the interaction between judged and presented frequency is a stable phenomenon. Such an experiment should be done with both intentional and incidental instructions, because the phenomenon may occur in one case but not in the other.

A second experiment which should be conducted is a continuous JOF test with lag as an independent variable. As mentioned at the beginning of this paper, continuous JOF paradigms have been used only occasionally and never with lag as a variable. Not only would such an experiment close an empirical gap but it would cast some more light upon hypotheses which explain the spacing effect in terms of encoding processes. For instance, Hintzman suggests that the locus of the effect lies with the weak encoding of the repetitions (as opposed to the initial presentations) at short lags. If this is so, then subjects on a continuous test should have lower performance with items repeated at short lags than with items repeated at long lags, i.e. their performance should parallel that for a terminal test and increase with lag. In all probability, though, such would not be the case as indicated by evidence from recognition studies which have employed the continuous paradigm. Shepherd and Teghtsoonian (1961) and Nickerson (1965) both found that recognition decreased with lag and one might reasonably expect the same to occur with judgments of frequency.

If such proves to be the case, how does one reconcile the apparent paradox of memory performance decreasing over lag with a continuous test but increasing over lag with a terminal test? One step towards the reconciliation seems to be a shift in the locus of the spacing effect from encoding to retrieval. This is implied in Figure 1 and the discussion of it in the Introduction. The overlapping list markers with repeated presentations at short lags are difficult to discriminate when retrieval is necessary. Time would be a factor here. Discrimination would not be difficult on a continuous test, if the testing came after only a few intervening items. This had been shown not only in the recognition studies mentioned above but also in paired-associate learning with a continuous paradigm by Peterson and his colleagues (e.g. Peterson, Wampler, Kirkpatrick, and Saltzman, 1963). However, the low discriminability of list markers encoded at short lags would have a detrimental effect if testing came many items later as, for example, in the terminal paradigm.

Secondly, if the Begg hypothesis proves to be substantiated by the suggested within-experiment comparison, then the locus of the effect lies with retrieval as opposed to encoding. This must be so since the subject during the test trial does not realize what the range of presentation frequencies will be and the interaction between actual and judged frequency cannot occur until the subject has formed a notion of this range. One is then faced with the theoretical problem of how the subject forms such a notion. It is possible to conceive of list markers of different strength with each repetition of an item encoding a list marker that is weaker than the preceding one. This would lead to smaller increments in judgments of frequency as presentation frequency increased (which is the general case) and also account for the finding of Hintzman et al. (1973) that the locus of the spacing effect is with the second of two presentations. However, such a view places the emphasis on encoding again and would not account for the effect which the presentation frequency range has upon JOF performance. The solution seems to lie under Howell's (1973) rubric of multiple-process in which notions of both strength and list markers are used. The latter are useful concepts in discussing the lag effect but not for explanations of the memorial representation of frequency. This seems to require a concept of derived strength and a major theoretical problem for students of event frequency will be the reconciliation of these two concepts into a meaningful unity.

REFERENCES

Anderson, J.R. & Bower, G.H. Recognition and retrieval processes in free recall. Psychological Review, 1972, 79, 97-123.

Begg, I. Estimation of word frequency in continuous and discrete tasks. Journal of Experimental Psychology (in press).

Begg, I. & Rowe, E.J. Continuous judgments of word frequency and familiarity. Journal of Experimental Psychology, 1972, 95, 48-54.

Craik, F.I.M. & Lockhart, R.S. Levels of processing: A framework for memory research. Journal of Verbal Learning and Verbal Behaviour, 1972, 11, 671-684.

D'Agostino, P.R. & DeRemer, P. Repetition effects as a function of rehearsal and encoding variability. Journal of Verbal Learning and Verbal Behaviour, 1973, 12, 108-113.

Gartman, L.M. & Johnson, N.F. Massed versus distributed repetition of homographs: A test of the differential-encoding hypothesis. Journal of Verbal Learning and Verbal Behaviour, 1972, 11, 801-808.

Greeno, J.G. Paired-associate learning with massed and spaced repetitions of items. Journal of Experimental Psychology, 1964, 67, 286-295.

Hintzman, D.L. Apparent frequency as a function of frequency and the spacing of repetitions. Journal of Experimental Psychology, 1969, 80, 139-145.

Hintzman, D.L. Theoretical implications of the spacing effect. In R.L. Solso (Ed.) Theories of Cognitive Psychology: The Loyola Symposium. Potomac, Maryland: Lawrence Erlbaum Associates, 1974.

Hintzman, D.L. & Block, R.A. Memory judgments and the effects of spacing. Journal of Verbal Learning and Verbal Behaviour, 1970, 9, 561-566.

Hintzman, D.L. & Block, R.A. Repetition and memory: Evidence for a multiple-trace hypothesis. Journal of Experimental Psychology, 1971, 88, 297-306.

Hintzman, D.L., Block, R.A., & Summers, J.J. Modality tags and memory ~~for~~ repetitions: Locus of the spacing effect. Journal of Verbal Learning and Verbal Behaviour, 1973, 12, 229-238.

Hintzman, D.L. & Rogers, M.K. Spacing effects in picture memory. Memory & Cognition, 1973, 1, 430-434.

Howell, W.C. Representation of frequency in memory. Psychological Bulletin, 1973, 80, 44-53. (a)

Howell, W.C. Storage of events and event frequencies: A comparison of two paradigms in memory. Journal of Experimental Psychology, 1973, 98, 260-263. (b)

Jenkins, J.J. Can we have a theory of meaningful memory?

In R.L. Solso (Ed.) Theories of Cognitive Psychology: The Loyola Symposium. Potomac, Maryland: Lawrence Erlbaum Associates, 1974.

Johnston, C.D. & Jenkins, J.J. Two more incidental tasks that differentially affect associative clustering in recall. Journal of Experimental Psychology, 1971, 89, 92-95.

Kintsch, W. Recognition learning as a function of the length of the retention interval and changes in retention interval. Journal of Mathematical Psychology, 1966, 3, 412-433.

Light, L.L. & Carter-Sobell, L. Effects of changed semantic context on recognition memory. Journal of Verbal Learning and Verbal Behaviour, 1970, 9, 1-11.

Madigan, S.A. Intraserial repetition and coding processes in free recall. Journal of Verbal Learning and Verbal Behaviour, 1969, 8, 828-835.

Melton, A.W. The situation with respect to the spacing of repetitions and memory. Journal of Verbal Learning and Verbal Behaviour, 1970, 9, 596-606.

Nickerson, R.S. Short-term memory for complex meaningful visual configurations: A demonstration of capacity. Canadian Journal of Psychology, 1965, 19, 155-160.

Peterson, L.R., Wampler, R., Kirkpatrick, M., & Saltzman, D. Effect of spacing presentations on retention of a paired-associate over short intervals. Journal of Experimental Psychology, 1963, 66, 206-209.

Rose, R.J. Spacing and presentation frequency effects in recognition and judgments of frequency. Unpublished research, 1973.

Rowe, E.J. Context effects in judgment of frequency. Bulletin of the Psychonomic Society, 1973, 4, 231-232. (a)

Rowe, E.J. Frequency judgments and recognition of homonyms. Journal of Verbal Learning and Verbal Behaviour, 1973, 12, 440-447. (b)

Rowe, E.J. Depth of processing in a judgment of frequency task. Unpublished research, 1974.

Rowe, E.J. & Rose, R.J. Instructional and spacing effects in judgment of frequency. Paper presented at the 35th Annual Meeting of the Canadian Psychological Association, Windsor, Ontario. (1974)

Rundus, D. Analysis of rehearsal processes in free recall. Journal of Experimental Psychology, 1971, 89, 63-77.

Shepherd, R.N. & Teghtsoonian, M. Retention of information under conditions approaching a steady state. Journal of Experimental Psychology, 1961, 62, 302-309.

Thorndike, E.L. & Lorge, L. The Teachers' Wordbook of 30,000 Words. New York: Teachers College, Bureau of Publications, 1944.

Underwood, B.J. Attributes of memory. Psychological Review, 1969, 76, 559-574. (a)

Underwood, B.J. Some correlates of item repetition in free-recall learning. Journal of Verbal Learning and Verbal Behaviour, 1969, 8, 83-94. (b)

Underwood, B.J., Zimmerman, J., & Freund, J.S. Retention of frequency information with observations on recognition and recall. Journal of Experimental Psychology, 1971, 87, 149-162.

Paper presented at the annual meeting
of the Canadian Psychological Association,
June, 1974

APPENDIX A

Instructional and Spacing Effects in Judgement of Frequency

Edward J. Rowe and Robert J. Rose

Memorial University of Newfoundland

The measurement of memory performance in tasks requiring subjects to judge how often a word occurred during a study trial has become increasingly common since the initial work of Hintzman in 1969. In the usual frequency judgement paradigm, subjects are presented with a long sequence of words which are repeated various numbers of times. Following the study trial, the subjects are then given a number of test words and are asked to judge how often these words appeared in the original list.

The experiment which I shall report examined the effect on frequency judgements of three independent variables: The number of repetitions of a word in the study list (i.e., its presented frequency), the spacing of repetitions (i.e., number of items intervening between successive presentations of each word), and the type of orienting instructions given to the subjects. Effects of the first two variables have already been well documented. As one would expect, judged frequency increases reliably with increases in presented frequency. The overall form of the relationship appears to be logarithmic, with subjects tending to overestimate the frequency of items presented relatively few times and underestimate the frequency of items presented many times in the list.

The effect of spacing of repetitions in memory experiments appears as an increase in memory for repeated items as the degree of spacing

APPENDIX A

between them increases. It is found with recall, recognition, and frequency judgement tasks involving not only words but nonsense syllables, sentences, and pictures as well. In a recent review of the spacing effect, presented at the 1973 Loyola Symposium, Hintzman points out that the peak or asymptote in performance usually occurs with a spacing of 15 sec between repetitions.

Our interest in this experiment focused primarily on the instructional variable and its relation to frequency and spacing. The subjects were given three types of instructions at the beginning of the experiment: intentional instructions, where they were fully informed of the nature of the ensuing frequency judgement test; nonspecific instructions, where they were told that a memory test would follow the list of words but its nature was not specified; and incidental instructions, where the subjects were told to rate each word on a dimension of strength as it was presented, with no mention being made of the frequency judgement test. Previous experiments have demonstrated frequency effects under both intentional and nonspecific instructional conditions. Howell, in a 1973 paper in J.E.P., showed that whether the subject is led to expect a frequency judgement or a recall test at the beginning of the experiment is irrelevant for subsequent frequency judgements. He therefore suggested that intentionality is not a critical factor as far as frequency judgements are concerned. However, both of Howell's conditions gave a set for some type of memory test, and the subjects probably engaged in active memory processing of the list. We were interested to see how frequency judgements obtained under truly incidental learning conditions would compare with intentional and nonspecific instructional sets.

With regard to the spacing variable, Hintzman has suggested that

APPENDIX A

the spacing effect might be attributable to involuntary rather than voluntary, or control, processes. The experiment tests this proposal by comparing the spacing effect under incidental and intentional conditions. If voluntary processing is not required for demonstration of the effect, it should obtain under conditions where no memory test is expected.

Let me briefly describe the design and procedure of the experiment. The words were of high Thorndike-Lorge frequency (A or AA) and had one or two syllables and five or six letters each. They were typed on 3 x 5 index cards and presented either two, three or five times, with each repetition being separated by 0, 1, 2, 4, 8 or 16 other items. This factorial combination produced 18 frequency x spacing cells. Five different words were allotted to each cell. These 90 words, their repetitions, and some necessary filler words, made up a basic pack of 327 cards. Six versions of the basic pack were devised, so that every word appeared twice at each frequency level and once at each level of lag.

The 144 subjects, all first-year Psychology students at Memorial University, were allocated equally to the six packs of cards within the three instructional conditions. As I have already stated, the subjects in the intentional condition were told to remember how often each word occurred in the pack of cards, subjects in the nonspecific condition were told to remember the words for an unspecified memory test, and subjects in the incidental condition were simply told to go through the pack of cards and rate the strength implied by each word on a seven-point scale. The presentation rate was paced by a metronome, which sounded every four seconds. Immediately after the subjects had gone through the cards they were given a test sheet containing

APPENDIX A

the 90 experimental words with 30 new words placed randomly among them, and were asked to write down how often each of the words appeared in the pack.

The mean frequency judgement for the words in each frequency x spacing cell, averaged across both words and subjects, is shown separately for each instructional condition in the first slide. Degree of spacing is plotted on the abscissa, and frequency of presentation is the parameter in each panel. The data were analyzed by an analysis of variance. As is indicated in the slide, there was a significant main effect of instructional set, with judgements increasing in order from the nonspecific to intentional to incidental conditions. The overall mean judgements (from top to bottom on the graph) were 2.80, 3.30, and 4.15. The three groups were all significantly different from each other. The main effect of frequency was also significant, and the means were ordered in the expected direction in each group. However, the size of the difference separating the three levels of frequency varied between the three groups, producing a significant instruction x frequency interaction. Note that words which occurred twice are overestimated and words occurring five times are underestimated in the nonspecific and intentional conditions, as is usually the case. Under incidental instructions, however, all three frequencies were overestimated, although the amount of overestimation for frequency 5 was slight.

The main effect of spacing was also significant but this factor can be evaluated more meaningfully by considering the two remaining significant interactions. First, spacing interacted with frequency, such that the spacing effect became larger as frequency increased. This was especially true of the intentional and incidental groups. Now consider

APPENDIX A

the interaction of spacing with instructions. It is apparent that the spacing effect is most pronounced in the incidental condition, but this is shown more clearly in the second slide. Here the data are collapsed across frequency. Not only is there a more rapid rise in the curve for the incidental group between a spacing of 0 and 2, but the function seems to continue to increase (at a reduced rate) over the higher degrees of spacing, whereas the curves for the other two conditions level off after four intervening items.

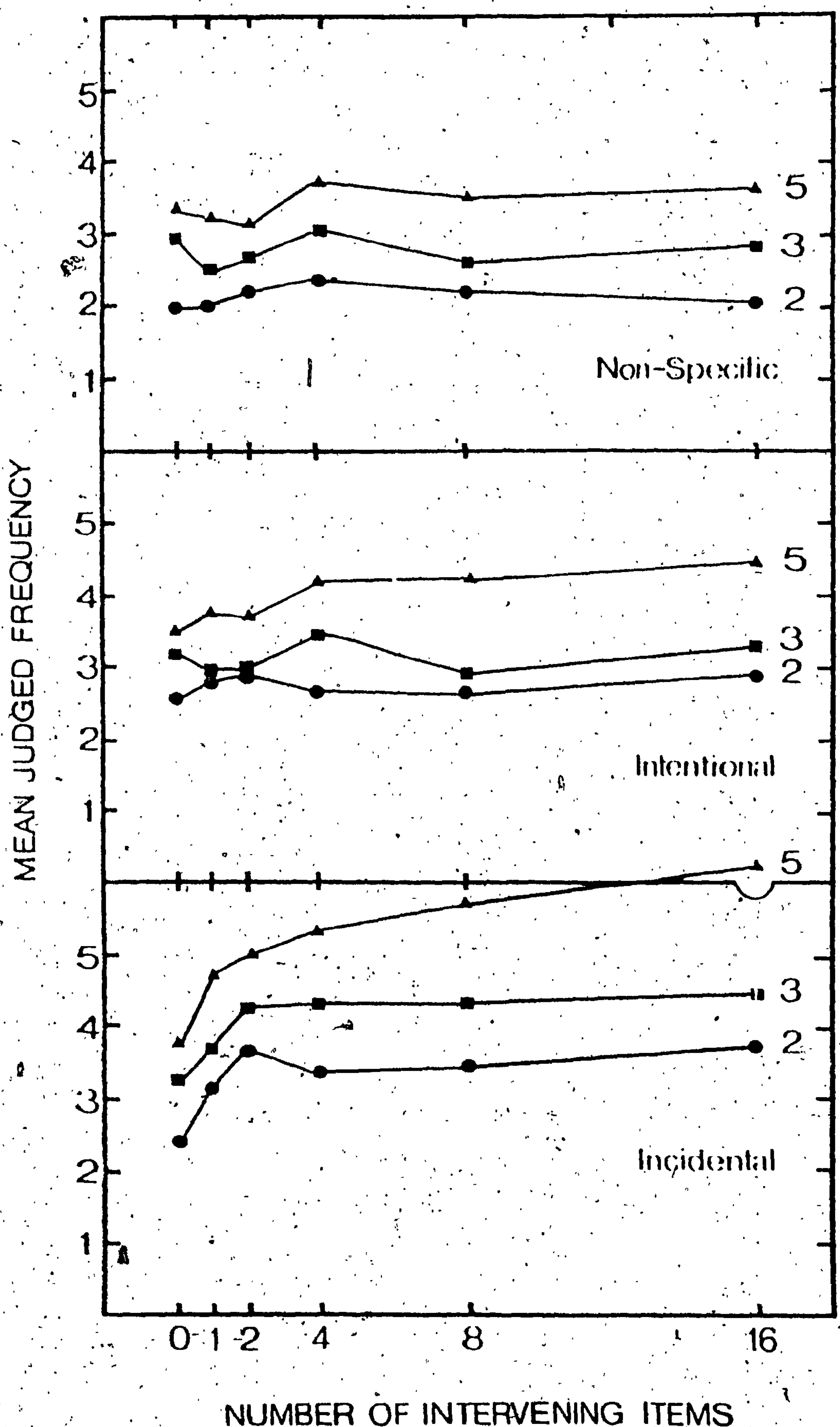
For the rest of the paper, I would like to comment on some aspects of the frequency and spacing effects. Contrary to Howell's (1973) results, we have shown that the type of instructional set does affect judgements of frequency. Intentional instructions produce higher judgements than nonspecific instructions and incidental instructions produce the highest judgements of all. Furthermore, the typical overestimation of low frequencies and underestimation of high frequencies is not found under incidental conditions, at least for the range of frequencies which we used here. I should point out that the inflated frequency judgements found with the incidental group are probably not due to some general response bias associated with the incidental condition, since, in a subsequent experiment, we have found that the degree of overestimation depends on the type of orienting instruction used. With a rating task similar to the one used here, where the subjects judged the goodness of the words instead of strength, the present findings were replicated almost exactly. On the other hand, a nonsemantic orienting task which required processing of the individual letters of each word gave results comparable to those obtained with nonspecific instructions here.

Interestingly, a group instructed to find a rhyme for each presented word performed very much like the rating group on the frequency judgement test. Thus we prefer to believe that the results obtained for the incidental group in the present experiment are not due to an overall tendency to give higher judgements. Rather, judging from the later experiment, the over-estimation of frequency seems to depend to some extent on the level of processing primed by the orienting task, in the sense used by Craik and Lockhart (1972).

Finally, the results are relevant to the theoretical explanation of the spacing effect, at least insofar as it applies to the frequency judgement task. Hintzman, in his 1973 review paper, distinguished between two classes of explanation: Those which depend on voluntary explanations attribute the effect to such things as differential rehearsal of, or differential attending to, or differences in the variability of encoding for, items which occur at low vs. high levels of spacing. Involuntary explanations stress the importance of consolidation and habituation processes. I will not discuss any of these in detail, but it should be obvious that our results favour the involuntary type of explanation. The subjects in the incidental group had no reason to treat the items repeated after a short interval any differently from the ones which had a larger spacing, yet not only was there a spacing effect in this condition, but it was enhanced compared to the conditions where a subsequent test was expected.

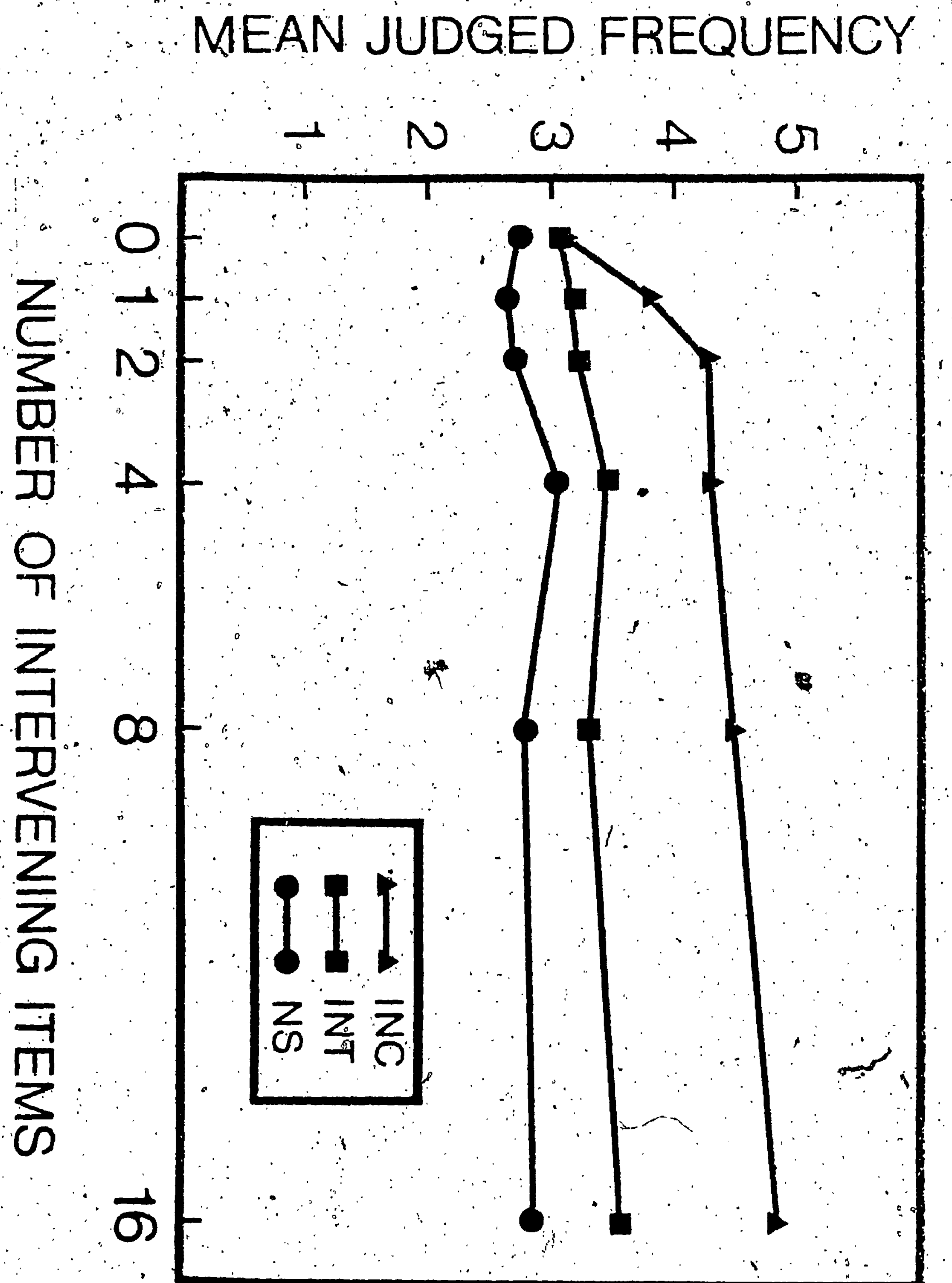
APPENDIX A

FIGURE 4. Mean judgments of frequency for each level of frequency as a function of spacing and group from the preliminary experiment by Rowe and Rose (1974).


MEAN JUDGED FREQUENCY
Non-Specific
Intentional
Incidental
5
3
2
0 1 2 4 8 16
NUMBER OF INTERVENING ITEMS

APPENDIX A

Figure 5. Mean Judgments of Frequency for Each Level of Spacing Collapsed Across Frequency


MEAN JUDGED FREQUENCY
1
2
3
4
5
NUMBER OF INTERVENING ITEMS
0
1
2
4
8
16
INC
INT
NS

APPENDIX B

INSTRUCTIONS

CONDITION I (EVALUATION)

You have a deck of cards with some words printed on them. I want you to go through the cards one at a time and indicate on the answer sheet your impression of how "good" each word is. For example, the word "movie" usually implies a fair degree of "goodness" for most persons because they associate it with enjoyable activity. On the other hand the word "exam" implies a fair degree of "badness" because people generally associate it with hard work, stress, anxiety, etc.

Indicate your feelings of "goodness" by writing a number from 1 to 7 opposite the number of the word. The number "1" indicates "very bad" and the number "7" indicates "very good" with number "4" indicating neutral feelings (i.e., "so-so"). In general, the stronger your feelings of goodness implied by a word, the higher will be the number you use.

You are to go through the pack, one card at a time, in time to the metronome. We are interested in immediate impressions so do not linger over any one word but keep in time to the "bleeps". Some of the words occur more than once but still give a number to every occurrence of every word. (Note that on the second page of the answer sheet you have to write your judgments to the left of the word number.)

APPENDIX B

INSTRUCTIONS

CONDITION II (CODING)

You have a deck of cards with some words printed on them. I want you to go through the cards one at a time in time to the metronome and decode the vowels as follows: a = 1; e = 2; i = 3; o = 2; u = 1. (I suggest that you write these values at the top of each answer sheet.) All other letters are to be ignored.

For example, the word "movie" would equal 7 (2+3+2), and the word "exam" would equal 3 (2+1).

To begin with you may find that you are rather slow. Do not worry about this; your speed will increase rapidly. It is very important that you keep in time with the metronome. If you have not finished with a word in the allotted time, turn over to the next card. There are many words from which we can obtain data and a few blanks near the beginning will not matter. Some of the words occur more than once but I still want you to give a number to every occurrence of every word.

(Note that on the second page of the answer sheet you have to write the values to the left of the word number.)

APPENDIX B

INSTRUCTIONS

CONDITION III (RHYMING)

You have a deck of cards with some words printed on them. I want you to go through the cards, one at a time, in time to the metronome and write on the answer sheet a word which rhymes with the word on the card. For example, a rhyme for the word "movie" might be "groovy" and for the word "exam" might be "ham". The rhymes are to be written opposite the appropriate word number on the answer sheet. (Note that on the second page of the answer sheet you have to write to the left of the word numbers.)

For some of these words, it will be difficult to think of a rhyme. If you cannot think of a rhyme by the time the metronome "bleeps", turn over to the next card. It is important that you keep in time with the metronome. (Do not turn over the cards too soon or too late.)

Some of the words occur more than once but I still want you to write a rhyme for every occurrence of every word. You may use the same rhyme for every occurrence if you wish.

(By the way, the rhymes may be proper nouns if you wish, e.g., the names of people, places, etc. Do not worry about spelling.)