LEVELS OF PROCESSING AND THE EFFECT OF SPACING UPON JUDGMENTS OF FREQUENCY

ROBERT JAMES ROSE

Levels of Processing and the Effect of Spacing Upon Judgments of Frequency

by

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

A Thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

Department of Psychology
Memorial University of Newfoundland
August, 1978

St. John's Newfoundland

Statement of Originality

This thesis is concerned with the "spacing effect", a phenomenon of memory. Basically the effect manifests itself as an increase in the probability of remembering a repeated item as the interval or spacing between the repetitions increases. The form of the effect is a function of a number of factors including number of repetitions, type of task performed on the learning trial, and length of retention interval.

The main contribution of the thesis to original knowledge lies in the finding that the general properties of the spacing effect can be accomodated by a levels of processing hypothesis. This hypothesis was initially proposed by Craik and Lockhart (1972) as a framework for human memory research and elaborated by Lockhart, Craik, and Jacoby (1976) to account for a number of memory phenomena. As applied to the effect of the spacing (i.e., the interval separating repetitions), it suggests that repetitions with short spacings are encoded easily by scanning recent memory for the previous occurrence of the stimulus whereas repetitions with long spacings initiate an attempt to reconstruct the original encoding of the stimulus. Those repetitions encoded by the scanning process are not encoded as deeply and hence not retained as well, as those encoded by the reconstructive process. The data were consonant with the levels of encoding hypothesis but they did not shed any light on the nature of the postulated scanning and reconstructive processes.

Abstract

The term "spacing effect" refers to the empirical fact that items which are repeated with few other items intervening between the repetitions are remembered worse than items which are repeated with relatively more interventions between the repetitions. The purpose of the set of spacing experiments reported here is to infer a cause of the spacing effect in a judgment of frequency paradigm by trying to discover conditions which will remove it.

Following Experiment 1, which compared continuous judgments of frequency of words made on the learning trial with terminal judgments of frequency made at the conclusion of the learning trial, it was inferred that the spacing effect arose from one or more of three possible origins: (1) a true memory deficit for massed repetitions together with a biassed tendency to overestimate the frequency of words repeated at non-zero spacings; (2) the use of different strategies on continuous and terminal judgments; (3) the deficient processing of repetitions at low values of spacing relative to repetitions at higher values.

Experiments 2 and 3 found evidence which was irreconcilable with the first two of these three possible origins of the spacing effect. Experiments 4 and 5 were

then carried out to test the relevance of the third possibility. Specifically, the levels of processing hypothesis was contrasted with the variable contextual encoding hypothesis. The former claims that the spacing effect arises because repetitions at long spacings receive deep, reconstructive processing while those at short spacings receive shallow scanning processing. The contextual encoding hypothesis, which enjoys some support elsewhere, attributes the spacing effect to the greater variability among the contents of the repetitions which occur at long spacings. The evidence from Experiments 4 and 5 generally supported the levels of processing hypothesis but in addition indicated that the encoding context is a factor which interacts with the measure of retention. Also the nature of the posited scanning and reconstructive processes still remains a mystery.

ACKNOWLEDGEMENTS

I should like to thank Ted Rowe, the supervisor of my thesis and its concomitant research, for his encouragement, guidance, and germane advice over the past four years. I should also like to thank the other members of my committee, Rita Anderson and Cathy Penney, for their prompt help and poignant criticisms. Finally, I should like to dedicate this thesis to my wife, Eeva Liisa, for her patient forbearance during my many years of part-time study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

This thesis contains a résumé of some research on the spacing effect together with a description of new research in the same area. The first section deals with a delineation of the spacing effect while the second section concerns theoretical explanations of the effect. The third section describes the research of the author, beginning with a parametric study which involved continuous and terminal judgments of the frequency of occurrence of words within a long sequence. The results of this experiment indicated that two processes could be relevant to the spacing effect: (1) the occurrence of different approaches to the continuous and terminal tests; (2) the deficient processing of repetitions at short spacings relative to repetitions at long spacings. The results also contradicted the finding from other research that the tendency for judgments to increase over non-zero spacings may be due to response bias. A further test of the effects of bias was carried out in Experiment 2 while an investigation of the possible use of differential test strategies was carried out in Experiment 3. Finally, Experiments 4 and 5 tested a levels of processing hypothesis which suggests that deficient processing is involved in the spacing effect.

Background

The term "spacing effect", also known as the "lag effect", refers to the empirical fact that memory for repeated items improves up to a point as the interval (i.e., the spacing or lag) between the repetitions increases. The spacing is usually measured in terms of the number of other items intervening between the repetitions, although some investigators have measured the spacing in units of time (e.g., Hintzman & Rogers, 1973). The facilitation of the retention of repeated items by increased spacing is exceedingly general and robust, having been found with a variety of materials and for several measures of retention (see the reviews by Hintzman, 1974; 1976).

In his two review articles, Hintzman distinguishes three varieties of spacing effect. The effect which he considers to be the most general is an increase in retention with spacing up to an interval of 15 seconds between repetitions, following which the retention curve asymptotes. Sometimes the retention curve drops slightly after reaching a peak (e.g., Peterson, Wampler, Kirkpatrick, & Saltzman, 1963; Madigan, 1969; Hintzman, Block & Summers, 1973). Hintzman considers this phenomenon to be an exception to the rule but Glenberg (1976) has shown that this decline is normal when the retention interval between an item's last appearance on the study trial and its appearance on the test is short.

A second type of spacing effect, which has been called the Melton effect, is a continued increase in the probability of retention beyond a spacing of 15 seconds (Melton, 1970). Hintzman distinguishes this version of the effect from the preceding one because, he claims, it is found only in a free recall task. However, we know now that such is not the case. Rose and Rowe (1976) showed that, in a frequency judgment task, judgments of frequency continued to increase up to spacings of at least one minute (16 intervening items at a rate of one item per 4 seconds) for higher levels of presentation frequency. This effect was especially marked for subjects required to carry out an incidental semantic task before making the judgments of frequency. Also, Paivio (1974, Experiment 2) found that recall of twice-presented concrete words increased over spacing up to a point and then levelled off, while recall of twice-presented pictures continued to increase over spacing, up to a lag of 48 items (240 seconds) at least. Taken together, these results indicate that the point along the spacing dimension at which the retention curve asymptotes is, to some extent at least, dependent upon task parameters and the type of material used. Finally, Glenberg (1976) found that where (or if) the spacing function attained an asymptote depends upon the retention interval. As mentioned in the previous paragraph, the function tends to peak and then decline when the retention interval is short. As the retention interval increases, the function tends to become flat or continues to rise slowly with high values of spacing. Thus the Melton effect turns out to be subsumed by general effect of spacing.

A third type of spacing effect referred to by Hintzman (1974) is a difference between massed presentations and distributed presentations, i.e., a difference between consecutive presentations and presentations with a spacing greater than zero (see also Underwood, 1970). With this phenomenon, which is called hereafter the massed presentation effect, there is no effect of spacing upon non-consecutive repetitions but the retention of massed presentations is depressed relative to the retention of the former. The evidence for distinguishing this type of spacing effect comes mainly from D'Agostino and De Remer (1972, 1973). In their 1972 paper, these investigators found that free recall of repeated sentences, for each of which students made up plausible short stories, showed the usual monotonic facilitation over spacing while cued recall following the same learning conditions showed only a massed presentations effect. D'Agostino and De Remer (1973) showed again a massed presentations effect (but no other effect of spacing) when subjects were required to recall repeated sentences to which they had formed, and overtly described, images. The conditions which produce a massed presentations effect are by no means clear and one may be tempted to dismiss the phenomenon as being a peculiar exception to the typical spacing effect. However, further evidence in support of the massed presentations effect comes from Hintzman (1969, Experiment 2), who tested words which had appeared twice in a study sequence at spacings of 0 to 16 intervening items. In a two-alternative forced-choice test, he found a preference to select, as being more frequent, words at non-zero spacings over massed repetitions but no preferences among the words at non-zero spacings. Thus the massed presentations effect may be more general than the indication given by the studies of D'Agostino and De Remer. This point will be taken up in Experiment 2 of this thesis.

In summary, then, there may be two separate phenomena involved in the spacing effect. One phenomenon, which will be labelled the spacing effect in this proposal, is an increase in retention over spacing up to some point. The position of this point along the spacing dimension depends upon several variables such as the presentation frequency, learning task, type of materials used, and the retention interval. The other phenomenon, the massed presentations effect, is a decrease in retention of consecutively repeated items compared to non-consecutive repetitions, among which no effect of spacing occurs. It should be made explicit that these phenomena are found with terminal testing paradigms, in which the subject receives a study or incidental learning trial and then is tested for retention. Occasionally, investigators of the spacing effect have used continuous testing paradigms whereby the subject responds to each item in the sequence on the initial trial. Such paradigms are not applicable, of course, to free recall but can be used with other measures of retention. The general finding from the few continuous studies of recognition is a monotonic decline in retention over spacing (Shepard & Teghtsoonian, 1961; Nickerson, 1965).

Theoretical Explanations of the Spacing Effect

The theoretical explanations reviewed by Hintzman (1974, 1976) will be retained here with the addition of one further explanation. Hintzman assumes that the spacing effect stems from a single source. This, of course, is not necessarily true, especially if the distinction of an independent massed presentations effect is valid. However, until further evidence for the independence of the massed presentations and the spacing effects is forthcoming, the assumption of a unitary effect of spacing arising from a single source will be retained.

The two main categories of explanatory hypotheses put forward by Hintzman (1976) are: (1) those theories which stress encoding variability, i.e., which attribute the spacing effect to the enhancement of items repeated at long lags due to the variability in the encoding of the repetitions of these items, and (2) those theories which stress deficient processing and subsequent poor retention of those items repeated at short lags. Within this second category are further divisions dependent upon the voluntary nature of the processes and the locus of their effect. Following the descriptions of the four hypotheses which Hintzman places under the rubric of deficient processing theories, mention will be made of a fifth hypothesis of the same category, namely the levels of processing hypothesis.

Encoding Variability Theories

Semantic Variability. The encoding variability hypothesis is generally ascribed to Martin (1968, 1972) who initially applied it to the area of paired-associate transfer. The semantic version of the hypothesis assumes that a given item can be encoded semantically in several different ways and that retrieval on a later test is enhanced as the number of different encodings given to an item on the study trial increases. Applying this view to the spacing of repeated items, it is further assumed that the encoding of a second presentation of an item (P_2) is more likely to be different from the encoding of the first presentation of that item (P_1) as the P_1-P_2 interval increases. Hence retention of a repeated item increases over spacing because the various repetitions of that item are more likely to be variably encoded as spacing increases.

A prediction which follows from the preceding hypothesis is that forcing subjects to encode all repetitions, regardless of spacing, in different ways should eliminate the spacing effect. This prediction has been tested and supported by at least three studies. Madigan (1969) found that differential semantic encoding of repeated words eliminated any effect of spacing when retention was measured by cued recall. D'Agostino and De Remer (1973, Experiment 2) also found no spacing effect when subjects were required to recall freely the object phrases of differentially encoded sentences following imagery instructions. Finally Gartman and Johnson (1972) found in a free recall study a large facilitatory effect of differentialy semantic encoding over similar semantic encoding, which simultaneously eliminated any effect of lag.

There is then some evidence from free recall studies to support the encoding variability hypothesis. However, that support is rather limited. For one thing, Madigan found that differential semantic encoding eliminated the spacing effect only for cued recall and not for free recall. Secondly, the aforementioned results of D'Agostino and De Remer (1973) are weakened by the fact that the corresponding control group, who formed images to identically repeated sentences, displayed an attenuated effect of spacing also, showing only a MP-DP effect.

Quite apart from these criticisms, there are implications of the theory which do not stand up to empirical test. For instance, Martin (1972, p. 65) suggests that an item will be encoded differently on P_2 than on P_1, if the subject fails to recognize P_2 as a repetition of P_1. This is more likely to happen if the P_1-P_2 interval is long than if it is short. Hence two predictions follow: (1) the recognition of repetitions decreases over spacing on the study trial, (2) the probability of retention of repeated items on a later test increases as the probability of recognition of their repetitions on the study trial decreases. The first prediction is true (Shepard & Teghtsoonian, 1961; Nickerson, 1965) but the latter is not (Bellezza, Winkler & Andrasik, 1975). In fact, Johnston and Uhl (1976) found over three levels of lag that only those P_2 words correctly recognized as repetitions showed a consistent spacing effect in a free recall paradigm. Thus correct recognition of an item as a repetition may even be a necessary condition for a spacing effect, which is completely contrary to Martin's view.

It is possible to argue that subjects still encode repetitions at long spacings differentially, even though the repetitions are correctly recognized. However, this argument is conditional upon support for the basic assumption of the semantic variability approach, namely that encoding an item semantically in different ways will enhance retention of that item on a later test. Such support has been found by Bevan, Dukes, and Avant (1966) for the recall of superordinates and by Gartman and Johnson (1972) who biassed the meanings of homographs by manipulating the two items which preceded them in the study sequence (e.g., meter-inch-foot and measure-yard-foot vs. meter-inch-foot and arm-leg-foot). On the other hand the assumption was refuted by Madigan (1969) in a free recall test, Schwartz (1975) in a paired-associate test, and Rowe (1973a, b) in a judgment of frequency test. Rowe found that forcing subjects to encode words in different semantic contexts lowered their judgments of the frequency of occurrence of those words whereas judgments of frequency, in common with other measures of retention, show the typical increase over spacing. Thus there is on balance no consistent support for the view that differential semantic encoding enhances retention and so this hypothesis cannot be seriously considered as a general explanation of the spacing effect.

Contextual Variability. A form of the encoding variability hypothesis which ascribes facilitation of the retention of repeated items to contextual, as opposed to semantic, encoding variability was proposed by Anderson and Bower (1972). According to their view, each time an item is presented on a study trial a bundle of contextual cues, called a list marker, becomes attached to the memorial representation of that item. Potentially, these list markers enable subjects to carry out several tasks such as to distinguish those items which appeared on a study trial from those which did not, to determine in which portion of a list an item appeared, to recall in which of several lists an item appeared, and to determine how often an item appeared in a list. As applied to the spacing effect, the argument is that the longer the spacing between repetitions, the more dissimilar will be the learning contexts in which the repetitions appear and the more distinct will be the list markers attached to the memorial representation of the repeated item. This increase in distinctiveness increases in turn the probability that the subject will retrieve at least one list marker (for recall or recognition) or all of the list markers (for a judgment of frequency) on a later test.

Glenberg (1976) also has put forward a type of encoding variability hypothesis, one which not only takes account of the contexts of the presentations of an item on the study trial but also refers to the context of the test. A visual analog of this view is shown in Figure 1. It is assumed that the encoded version of a nominal stimulus depends upon both that nominal stimulus and the context in which it appears. Furthermore the context changes in an orderly fashion over time, i.e., to quote Glenberg, the

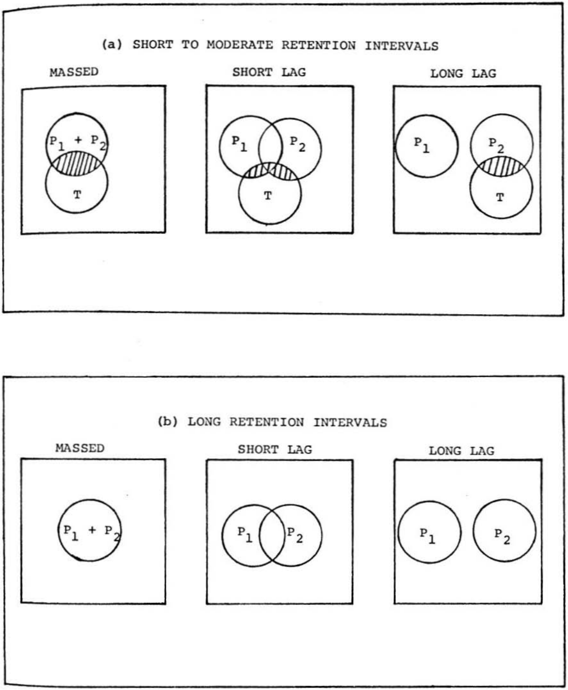

Figure 1: Visual analog of the effects of contextual variability upon retrieval of a repeated item from memory. (After Glenberg, 1976, Figure 2.)

"context at time n+2 is more similar to the context at time n+1 than to the context at time n" (p. 13). This orderly change is depicted in Figure 1 by the gradual separation of P_1 and P_2 as lag increases.

When the subject is tested for a repeated item, he/she will be successful in retrieving it to the extent that the functional stimulus at the time of the test (T) matches the encoded versions of P_1 and P_2. At short to moderate retention intervals, T should overlap with P_2. If the spacing between P_1 and P_2 is also short to moderate, T will overlap with P_1 as well. The degree of total overlap determines the probability of retention and is shown by the cross-hatched areas in Figure 1(a). One can easily see how this model predicts an inverted-U function for short retention intervals as spacing increases from zero. On the other hand, when the retention interval is long, T is deemed to be only weakly related to P_1 and P_2 and does not re-instate one more than the other. Here retention will be poorer than at short to moderate retention intervals but the spacing function will continue to rise slowly or flatten out.

The strength of Glenberg's view lies not merely in accounting for certain empirical phenomena but also in shifting theoretical emphasis from the fluctuating state of an encoded stimulus (between availability and unavailability) to the fluctuating state of the experimental context. In this he provides a link with the work of Tulving and his associates (Thomson & Tulving, 1970; Tulving & Thomson, 1973) who stress the role of context regarding the availability of encoded stimuli at the time of testing. However, a major problem has been to discover what constitutes a change in context. As with the differential semantic encoding hypothesis, one would predict that induced encoding variability (in a contextual sense this time) should eliminate the spacing effect. Contrary to these expectations, changing the input modality from visual for P_1 to auditory for P_2 (or vice versa) does not attenuate the spacing effect (Hintzman et al., 1973; Wells & Kirsner, 1974). Likewise presenting P_1 visually with silence and P_2 visually but accompanied by a tone does not alter the effect of spacing (Hintzman, Summers, Eki, & Moore, 1975), nor does requiring subjects to carry out different semantic rating tasks on P_1 and P_2 (Shaughnessy, 1976). The weakness of the contextual encoding variability hypothesis then is not so much that there is evidence to refute it as that no evidence to date supports it. However, the concept of "context" is very broad and the possibility remains that some sort of induced variability of context can be shown to eliminate the spacing effect.

Deficient Processing Theories

Theories falling under the rubric of deficient processing can be further subdivided according to the voluntary versus involuntary nature of the process and the locus of their effect. When this is done the following table emerges (see Hintzman, 1976):

Locus of Processing	Voluntary Processing	Involuntary Processing
Between P_1 and P_2	Rehearsal Theory	Consolidation Theory
During P_2	Attention Theory	Habituation Theory

These theories will now be examined in turn, beginning with those which postulate the locus of processing as being between P_1 and P_2. Then a fifth explanation, the levels of processing hypothesis, will be discussed. This hypothesis places the locus of the relevant processing during P_2 but could involve either voluntary or involuntary processing.

Rehearsal Theory. This view suggests that the spacing effect occurs because items presented at long lags receive more total rehearsals than items presented at very short lags. Rundus (1971) found support for this hypothesis using a free-recall paradigm during which subjects were instructed to rehearse aloud. However, there is considerable evidence which is difficult to reconcile with the rehearsal theory, in particular the persistence of an effect of spacing with incidental learning, where subjects have no reason to rehearse (Rose & Rowe, 1976). Thus the rehearsal theory at best applies only to situations of intentional learning and cannot be considered as a general explanation of the spacing effect.

Consolidation Theory. The consolidation theory states basically that time is required for a memory trace to "consolidate", i.e., to reach a state of relative permanence or, in other words, to be transferred into long-term memory (LTM). The source of the consolidation or transfer process is assumed to be the short-term memory (STM) trace, laid down immediately after the occurrence of the stimulus event. If certain other events occur during the period of consolidation of a memory trace, then the transfer of LTM will be incomplete and the probability of retrieval of that trace will be low.

One event which is assumed to disrupt consolidation is electroconvulsive shock. Indeed most of the evidence in support of the consolidation theory comes from animal studies which have shown that the delivery of electroconvulsive shock shortly after a stimulus event disrupts memory for that event. With reference to the spacing effect, one version of the theory assumes that P_2 of an item is analogous to the electroconvulsive shock is the sense that P_2 attenuates the consolidation of P_1 if it closely follows P_1. Hence the long-term memory of items repeated at short lags is poor.

A difficulty for this formulation of consolidation theory is the evidence of Hintzman et al. (1973). They "tagged" the two presentations of repeated words by using different input modalities and found that the second presentation was remembered worse than the first when the spacing between the repetitions was short. Assuming that P_2 is analogous to electroconvulsive shock predicts on the other hand that P_1 would be remembered worse than P_2.

There are, however, other versions of consolidation theory which do not draw an analogy between P_2 and electroconvulsive shock. For instance, Landauer (1969) suggests that a response to a stimulus event generates neural activity which is at a maximum just after that event and decays monotonically during the consequent periods (see Figure 2). This neural activity represents the process of consolidation. A reinforcing event, such as a second presentation of the stimulus event, has a probability of re-instating the original neural activity but only up to the maximum. The probability of recalling the stimulus event later from LTM depends upon the total neural activity occurring during the existence of the STM trace, i.e., upon the area beneath the curve(s) in Figure 2. For a twice-presented item, this area is a maximum when a repetition occurs following the termination of the neural activity associated with the preceding presentation of that event.

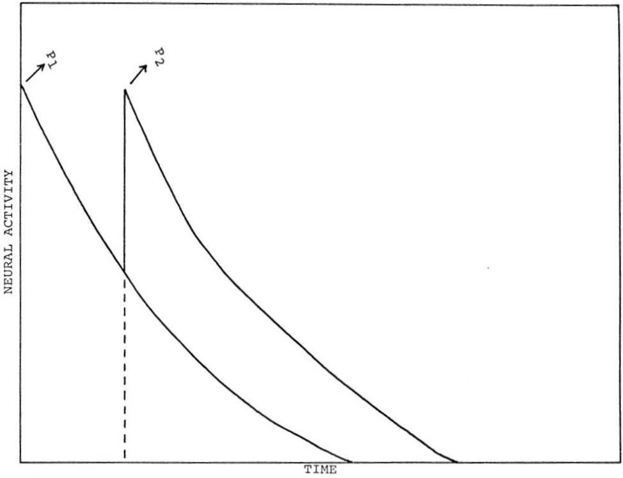

Figure 2. Schematic representation of the neural activity following two presentations of an item. (Adapted from Landauer, 1969.)

The preceding version of the consolidation theory, then, can account for P_2 being weaker than P_1 when the intervening spacing is short, if one represents the trace of P_2 by the increment to the curve representing P_1. However, there are other problems for this version of the consolidation theory. For instance, the theory predicts that the spacing effect on a later test is dependent only upon the time separating repetitions and not upon the number of intervening items. Hintzman and Rogers (1973) did find that the inter-repetition interval was a major factor but they also found that filling the interval with other items, as opposed to leaving it blank, increased judgments of frequency.

Another prediction of the theory is that retention of a repeated item on a later test will reach an asymptotic value with increased spacing, with the asymptote being reached when the consolidation of an item is finished. Two types of evidence go against this. One is the interaction between the spacing function and retention interval (Glenberg, 1976), i.e., the tendency for the curve to continue to rise over lag as the retention interval increases. The second is the interaction between presentation frequency and spacing, such that the asymptotic value increases as frequency increases (Underwood, 1969, 1970; Rose & Rowe, 1976). To accomodate this finding, each repetition must take longer to consolidate than the preceding presentation of that item. In summary, then, the consolidation hypothesis would have great difficulty in accounting for the changes in the asymptotic value.

No current formulation of the consolidation hypothesis is consonant with all of the empirical data. Before turning to other hypotheses, though, mention should be made of two other views related to this sub-section. One is the suggestion by Glanzer (1969) that the spacing effect is due to the limited capacity of a short-term memory store. Specifically, the two presentations of a repeated item have a less than additive effect if P_2 occurs while P_1 is still in STM. This view is the same as Landauer's (1969) with the concept of STM substituting for the concept of consolidation. The only difference is Glanzer's implication that the spacing effect is measured over number of intervening items instead of elapsed time.

The second view is the dual trace consolidation hypothesis of Wickelgren and Berian (1971). They postulate that potential STM and LTM traces are acquired during the period of active study and that consolidation converts each potential trace into a retrievable trace. Consolidation of the STM trace is very rapid and is followed by rapid monotonic decay. Consolidation of the LTM trace begins only 10 seconds or so after the termination of active study and requires about 10 or more seconds to be complete. A LTM trace may persist for years and the total memory strength is the sum of the STM and LTM components at the time of retrieval. Unfortunately, the theory is not sufficiently precise to allow one to make clear predictions concerning the effects of repetition. At the very least though, the view indicates that the maximum value of a retention function over lag should occur about 20 seconds following the termination of active study. In that case, the dual trace consolidation theory suffers from the same drawbacks as does Landauer's.

Habituation Theory. The habituation-recovery hypothesis, first suggested by Hintzman (1974), states that when P_1 occurs, its memorial representation habituates (i.e., enters a state of adaptation or refractoriness). This state of habituation gradually disappears, but if P_2 occurs before recovery is complete, then P_2 will not be encoded at full strength. Thus the second of two presentations of an item will be encoded deficiently, if the spacing between the presentations is short, and this in turn accounts for the spacing effect.

There appear to be two basic weaknesses with the habituation-recovery hypothesis. First, it is similar to Landauer's consolidation hypothesis in that it predicts that the probability of retention of a repeated item will reach an asymptotic value at some fixed value of the temporal spacing between the repetitions. As such, habituation theory suffers from the same criticisms as consolidation theory. The second weakness is that there is no direct evidence to clearly support the habituation-recovery hypothesis. For instance, Hintzman (1974) argues that rehearsal of an item should prolong the habituation process. In this case subjects who rehearse should reach the asymptotic value at a higher level of lag than subjects who do not rehearse, but Rose and Rowe (1976, Experiment 1) found the opposite. Hintzman, Summers, and Block (1975b) argued similarly that increasing the exposure duration of P_1 should increase the subsequent period of recovery from habituation and hence prolong the effect of spacing. They found, however, that the form of the spacing function was not altered by manipulating the period of duration of P_1. The habituation-recovery hypothesis appears then to be a weak contender as an explanation of the spacing effect.

Attention Hypothesis. According to the attention hypothesis, subjects pay less attention to P_2 when it closely follows P_1 than when it does not. Shaughnessy, Zimmerman, and Underwood (1972) supported this view when they found that subjects in a self-paced free recall experiment spent less time studying massed repetitions than spaced ones. Similar support comes from Johnston and Uhl (1976), whose subjects had to respond to a weak auditory signal as a secondary task, while being primarily engaged in studying a list of words for free recall.

Reaction time (RT) to the signals was measured. It was assumed that latency to the auditory stimulus varied inversely with the amount of attention subjects were paying to the secondary task and hence directly with the effort they were making in studying the words. Mean RT decreased monotonically over four repetitions for massed items but increased somewhat over four repetitions for distributed items. Thus the subjects appeared to be spending less effort in studying the repetitions of massed items than in studying the repetitions of spaced items.

Further support comes from Elmes, Greener, and Wilkinson (1972) who found that words which followed massed pairs of presentations were better recalled than words which followed distributed repetitions. This result suggests that subjects "relax" somewhat during massed repetitions and then apply more effort to the following word. Also, a study by Zimmerman (1975), as re-interpreted by Hintzman (1976), found that probability of recall varied directly with effective study time, regardless of lag and presentation frequency. In other words, massed items were poorly recalled because they received relatively little effective study time.

There is some evidence then to support the hypothesis that the spacing effect arises from differential attention to massed and distributed items. The finding by Johnston and Uhl (1976), that the difference in effort paid to massed and distributed items increases over presentation frequency, is especially interesting as a possible explanation of the interaction between spacing and frequency, i.e., the greater effect of spacing found with higher levels of presentation frequency. On the other hand Hintzman, Summers, Eki, and Moore (1975, Experiment 3) found that the number of eye fixations given a picture was independent of spacing and dropped over number of presentations for both massed and distributed items. This may indicate that either RT in the secondary task by Johnston and Uhl or the number of eye fixations (or both) is a poor indicator of attention.

A weakness of the attention hypothesis, as pointed out by Hintzman (1976), lies in the correlational nature of the supporting evidence, which is no substitute for evidence that the manipulation of attention can eliminate the effect of spacing. Hintzman et al. (1975b) found that neither manipulation of monetary incentive nor overt, as opposed to silent, rehearsal affected the shape of the spacing function. Thus, there is no evidence that manipulating subjects' attention will eliminate the spacing effect. One of the problems with such manipulations, however, lies in deciding what constitutes a dimension of attention. An alternative view which attempts to obviate this problem will be discussed in the next section.

The Levels of Processing Hypothesis. A hypothesis attributing the spacing effect to differential levels of processing is a fifth type of deficient processing theory. The view stems from the arguments of Craik and Lockhart (1972) that the long-term retention of an item is enhanced if that item is processed to a "deep", semantic level. The application of this theoretical approach to the effects of spacing has been made explicitly by Lockhart (Note 1), Lockhart, Craik, and Jacoby (1976), and Rose and Rowe (1976).

Briefly, the explanation states that, when an item is repeated in a study sequence, the subject attempts to contact the memory trace of the first presentation. When repetitions are close together, this contact is made relatively easily by scanning recent episodic memory. As the spacing between repetitions increases, more effort than mere scanning must be employed and the subject must then reconstruct something approaching the encoding of the original event. This reconstruction process would involve a deeper level of processing than the scanning process and hence would lead to better long-term retention. It should be made clear, however, that recognition within the study sequence of a repetition as such decreases over lag (Shepard & Teghtsoonian, 1961; Nickerson, 1965). Hence the increase in long-term retention over lag will reach an asymptote when the facilitating effect of deep processing is offset by the decreased probability of recognition of repetitions.

The levels of processing hypothesis has some explanatory advantages over other hypotheses concerned with the spacing effect. For instance, it can account for the data which support the encoding variability hypothesis by arguing that altering the encoding context from one presentation of an item to the next forces subjects to use a deeper retrieval process than the scanning process. At the same time, the levels of processing hypothesis accounts for the empirical fact that probability of long-term recall of an item increases as probability of recognition of its repetitions on the study trial increases (Melton, 1967; Bellezza et al, 1975; Johnston & Uhl, 1976). Note that the semantic version of the encoding variability hypothesis implies the opposite prediction.

The levels of processing hypothesis can also account for the counter-intuitive findings of Bjork and Allen (1970), Robbins and Wise (1972), and Tzeng (1973) who found that recall of twice-presented items was better when a difficult task came between the repetitions than when an easy task was interpolated between P_1 and P_2. These investigators all suggest that their results support an encoding variability hypothesis. However, one could also argue that a difficult task interpolated between repetitions necessitates the use of the reconstructive process when subjects attempt to retrieve the trace of the first presentation of an item.

Evidence favouring a levels of processing hypothesis over an attention hypothesis arises from the finding of Lockhart (Note 1). As Hintzman (1976) points out, the voluntary attention hypothesis implies "that processing effort can be allocated among stimuli in a flexible way," which in turn implies that items at long lags are retained well at the expense of items at short lags. Hence, if the spacing between repetitions is kept uniform within a study list, with spacing manipulated between lists, then there should be no effect of spacing unless repetitions of items at long lags are processed differentially at the expense of once-presented items. Lockhart found that recall of repeated words increased over spacing, even though spacing varied as a between-list factor. In addition, he found no differences across lists in the recall of single items. This finding of a benefit with spaced repetitions without a corresponding decrement in the retention of single items is difficult to reconcile with the voluntary attention hypothesis but is easily accounted for by a levels of processing approach.

The levels of processing explanation then is consonant with several empirical findings and has certain advantages over the hypotheses described above. However, there are two points to be made here. First, the hypothesis will have to be expanded to account for the tendency for retention of a repeated item to continue to improve over spacing as the retention interval increases. As was mentioned above, the levels of processing view holds that an asymptote is reached when the facilitating effect of deep processing of P_2 is offset by the decreased probability of recognition of P_2 as a repetition (see Lockhart, Note 1). This view is independent of the interval of time between P_2 and the final test presentation and is therefore not consonant with the findings of Glenberg (1976). While Lockhart et al. (1976, p. 77) do indicate the importance of the similarity of presentation and test encodings, they have not yet incorporated this notion into their discussion of the spacing effect.

The second point is that the levels of processing hypothesis has received little direct experimental testing. One exception is a study by Shaughnessy (1976) who asked subjects to carry out different rating tasks on the two or three presentations of each word. This procedure was based on the assumption that different rating tasks would require the subjects to pay closer attention to each presentation of an item (or in other words process each presentation to a deep level) and thus eliminate any effect of spacing. As it turned out, the spacing effect was not eliminated. However, there were certain methodological problems with this study (mentioned by Shaughnessy himself) and a further test of the levels of processing approach will be suggested later as part of the research proposed in this thesis.

THE RESEARCH

Introduction to the Research

The research described in the following sections of this thesis was conducted using judgment of frequency tests as the primary measures of retention. Since the spacing effect is a phenomenon of repetition, the judgment of frequency test seemed to be particularly suitable because it is a direct measure of memory for repetitions. In addition, the judgment task is useful because it lends itself readily to both continuous testing, where the subjects judge the frequency of occurrence of each item as they come to it on the study trial, and terminal testing, where the subjects judge the frequency of occurrence of each item after they have been through the study sequence. Performance on the continuous test may be taken as a measure of encoding of each item as a repetition while performance on a terminal test may be taken as a measure of long-term retention.

The research encompassed five experiments. In Experiment 1, subjects were given a continuous judgment task followed by a terminal test. In other words, they went first through a long sequence of words and judged how often each word had been presented in the sequence up to the current occurrence. Following this task, they were given a list of words and asked to judge how often each of these items appeared in the study sequence. Experiment 1, then, involved a comparison between continuous and terminal judgments of frequency to determine whether the continuous task on the study trial eliminated the effect of spacing on the terminal test.

The results of Experiment 1 gave rise to further questions. One was whether the massed presentations effect mentioned in the first section of this paper was valid. The evidence bearing on this question involved a non-biassed measure, the discrimination coefficient, which indicated that a distinction between repetitions at a lag of zero and repetitions at non-zero lags was not warranted, contrary to the results of Hintzman (1969). A second experiment was therefore carried out to examine this distinction further. A second question arising from the initial experiment was whether subjects employ with terminal judgment tasks some strategy which they do not use with continuous judgment tasks. This possibility was examined in Experiment 3.

The combined results from the first three experiments indicated that the decline in performance on the terminal judgments of frequency arose from forgetting over the relatively long retention intervals which were involved in the terminal tasks. The massed items in particular showed the effects of forgetting, a finding which indicated that they were processed deficiently relative to items repeated at longer lags. This argument was consistent with a type of levels of processing hypothesis which was tested more directly in Experiments 4 and 5.

Experiment 1

The first experiment had two main purposes: (1) to determine if a continuous judgment of frequency test on the study trial would eliminate the effect of spacing on the terminal test, and (2) to determine if the factor of spacing had an effect on continuous judgments of frequency. In addition, continuous and terminal judgments of frequency could be compared at various levels of spacing. These purposes will be discussed in order.

As to the first purpose, if the continuous task eliminates the spacing effect, we have now established a limit to the phenomenon. Further, and more importantly, such an outcome would suggest the possibility that the source of the spacing effect was some inefficient "control" process. In other words, subjects who do not carry out a continuous judgment task on the study trial may use a strategy which is different and inefficient compared to that used by subjects who do carry out a continuous task. This inefficient strategy would lead in turn to an effect of spacing on the terminal judgment of frequency test.

The use of a continuous judgment of frequency test as a measure of retention is comparatively rare among reported studies. At the time when Experiment 1 was undertaken it had been used only twice (Begg & Rowe, 1972; Begg, 1974) and only once was it followed by a terminal test (Begg, 1974). In neither of these studies was spacing a variable because the lag between repetitions was randomized. Thus a major concern of the first experiment was to discover whether the spacing effect on the terminal test would be altered by instructing the subjects to carry out on the study trial the precise task required by the terminal test. Although Rose and Rowe (1976, Experiment 1) used subjects who knew the nature of the terminal test and still showed the usual effect of spacing, it is possible that their covert strategies were inefficient relative to the process involved in an overt continuous judgment of frequency task.

There are two possibilities regarding the effect of spacing upon the terminal test in Experiment 1. One is that there will be the typical spacing function described in the introduction to this thesis, since this has been found in experiments where the subjects have been informed before the study that they would be tested for judgments of frequency (e.g., Rose & Rowe, 1976, Experiment 1).

The second possibility is that there will be no effect of lag on the terminal test. This arises from the studies of Begg and Rowe (1972) and Begg (1974), who found that subjects were very accurate at making continuous judgments of frequency of items which appeared up to 17 times. As an explanation of their results, Begg and Rowe (1972) speculated that subjects carrying out a continuous test formed paired-associates with the word being the stimulus term and its current frequency being the response term. Given that this is the case, the empirical results indicate that subjects are very accurate at encoding these paired-associates. Hence one would expect that memory for the frequency of occurrence of an item would deteriorate only over the retention interval between final presentation on the study trial and presentation on the test trial, i.e., the terminal judgments would be independent of spacing. One might of course reconcile the suggestion of Begg and Rowe with the typical spacing effect by arguing that paired-associates formed with items appearing at small lags are unstable and hence not recalled accurately on the terminal test. This argument merely predicts, however, that items repeated at small values of spacing are given inaccurate judgments relative to items repeated at large values of spacing. It does not predict necessarily lower judgments, which is certainly the case when a terminal judgment of frequency test is not preceded by a continuous test. In any event, one can predict from the speculations of Begg and Rowe (1972) that there should be no effect of spacing upon the mean values of the terminal judgments in Experiment 1.

With regard to the effect of spacing upon the continuous judgments, there are three possible outcomes. First, judgments of frequency could increase with spacing because they do so on a terminal test and one might expect continuous and terminal tests to produce similar results, since they involve the same measure of retention. On the other hand, there could be no effect of lag on a continuous test, based upon the arguments derived from Begg and Rowe (1972) that subjects making continuous judgments form accurate paired-associates. Thirdly, continuous judgments could decrease over spacing. Such a decrease is found in continuous recognition (Shepard & Teghtsoonian, 1961; Nickerson, 1965) and one might expect that the two measures of retention would resemble each other on continuous tests as well as on terminal tests.

Method

Materials. The materials were the same as those used by Rose and Rowe (1976). The subjects saw a list of 90 experimental words, each of which appeared 2, 3, or 5 times with spacings of 0, 1, 2, 4, 8, or 16. In addition, there were 27 filler items appearing once each. This design yielded 18 frequency x spacing cells with five experimental words allotted to each cell.

The items were common words (rating of A or AA) according to Thorndike & Lorge, 1944) of one or two syllables and five or six letters each. They were typed in lower case on white index cards for presentation to the subjects. The pack of cards consisted of five sections such that, with one exception, one word from each cell, together with its repetitions, occurred in each section. The exception was the word occurring five times with a lag of 16, whose final repetition always occurred outside its allotted section. Within each section, the 18 experimental words were allotted random positions within the constraints of the lag variable. In addition, the items appearing at short lags generally occurred between the repetitions of items occurring at longer lags, in order to keep the overall length of the sequence to a manageable size. Eight of the 27 filler items served as primacy and recency buffers, four items in each. Altogether there were 327 cards in a pack.

In order to counterbalance for specific-item effects, six packs of cards were constructed such that, across the packs, every word appeared once at each level of spacing and twice at each level of frequency. Once the first pack was constructed, the remaining five packs were derived from it by keeping the frequency x lag cells constant and rotating the sets of five words which occupied these cells. The fillers were constant across all packs.

Subjects. The subjects were 24 undergraduates of Memorial University who were paid $2.00 each for their participation. Four subjects were assigned to each of the six packs of cards. The testing was carried out in small groups of not more than six subjects each.

Procedure. The subjects were instructed to go through the pack of cards at their own pace and to write in the appropriate space on the response sheet the frequency of occurrence of each word within the sequence, up to and including the current presentation. Thus, the first time they met a word they wrote "1", the second time they met the same word they wrote "2", etc. The subjects had a dozen practice words and were allowed to ask questions before beginning the actual experiment. Half of the subjects went through their packs in reverse order.

When they had finished the continuous task, the subjects were given an unexpected terminal test. The test sheet contained all 90 experimental words and 30 new words. The filler words were not tested. The subjects were asked to judge how often each word had appeared in the sequence, assigning zero to the new words. The terminal test was also unpaced. The entire experiment took about 35 minutes.

Results

The mean judgment of frequency on the terminal test and the mean final judgment on the continuous test were found for each presentation frequency x lag cell for each subject. The overall means (i.e., averaged across subjects) are given in Table 1 and depicted in Figure 3. To begin with, separate analyses of variance were carried out on the continuous and terminal test data and summaries of these are reported in Tables A and B of Appendix A. In this and subsequent experiments, a Newman-Keuls test was carried out on individual means when justified by a significant main effect. Unless otherwise indicated, all significant effects from the analyses of variance reported in the results sections had p values of less than .001 and all significant differences between individual means had p values of .05 or lower. (All reported summaries of analyses of variance from this and subsequent experiments will appear in Appendix A). The results will be presented in an order which reflects the order of the purposes of the experiment, beginning first with the terminal judgments.

Terminal Judgments. The analysis of the terminal test data found significant effects of presentation frequency, $F(2,46) = 75.21$, lag, $F(5,115) = 15.23$ and frequency x lag, $F(10,230) = 4.70$. Because the interaction was significant, one-way analyses of variance were performed for each level of frequency. These showed no effect of spacing at a frequency of 2, ($F < 1$) but significant effects at frequences of 3 and 5, $F(5,230) = 3.82$ and 21.30 respectively, p's $< .01$. For a frequency of 3, the mean judgment at spacing of 0 was significantly lower than the mean judgments at all other values of spacing, except for lag of 1. For words presented 5 times,

TABLE 1

Mean Judgments of Frequency by Type of Test
(Experiment 1)

	Spacing						
Frequency	0	1	2	4	8	16	$\overline{X}$
			Terminal Test				
2	2.19	2.40	2.33	2.48	2.36	2.58	2.39
3	2.48	2.68	3.14	3.05	3.03	3.19	2.93
5	2.69	3.63	3.33	4.06	4.13	4.58	3.74
$\overline{X}$	2.46	2.91	2.94	3.19	3.17	3.45	3.02
			Continuous Test				
2	2.18	2.29	2.28	2.30	2.38	2.37	2.30
3	3.13	3.08	3.28	3.21	3.28	3.40	3.23
5	5.13	4.96	4.73	4.95	5.08	5.64	5.08
$\overline{X}$	3.48	3.44	3.43	3.49	3.58	3.80	3.54

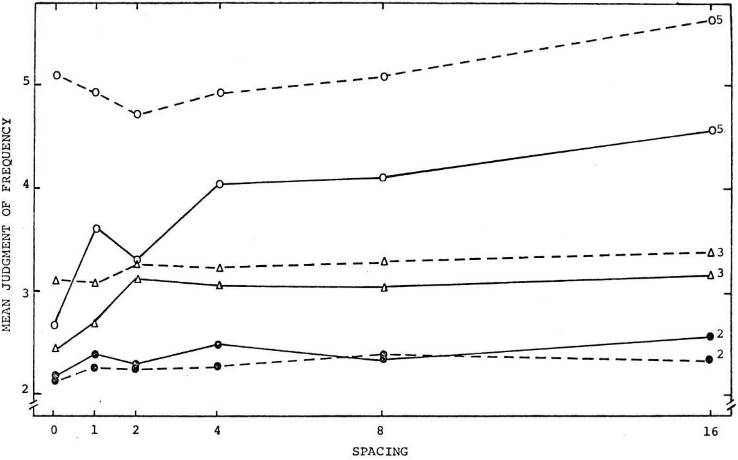

Figure 3. Mean judgment of frequency as a function of spacing and presentation frequency (2, 3, 5) for the continuous (---) and terminal (——) tests (Experiment 1).

the mean judgment at lag of 0 was significantly lower than all the others while the mean judgment at lag of 16 was significantly higher than all others.

Curvilinear regression analyses were also carried out at each level of frequency. There was no significant regression of judgment of frequency upon spacing at frequency of 2. For thrice-presented words, both the linear and cubic components were significant, F's(1,230) = 8.47 and 4.97 respectively, p's $<$.05. At a frequency of 5, the linear, quadratic, and cubic components were all significant, F's(1,115) = 75.79, 12.64, and 7.41 respectively, p's $<$.01.

Continuous Judgments. The analysis of the final judgments on the continuous test yielded significant effects of frequency, $F(2,46) = 1072.72$ and spacing, $F(5,115) = 4.55$ as well as the frequency x spacing interaction, $F(10,230) =$ 3.42. Once again, one-way analyses of variance were performed for each level of frequency. This yielded a significant effect of lag only at a frequency of 5, $F(5,230) =$ 12.02, which was due to the mean judgment at lag of 16 exceeding the mean judgments at all other values of spacing.

Judgments of Frequency by Word Position. The judged frequencies of each word within each of the five sections of the sequence were averaged across subjects for each frequency x spacing cell. These data were plotted and inspected to determine whether it was affected by a change in judgmental criteria as subjects progressed through the sequence.

Accordingly, graphs similar to Figure 3 were examined for each of the frequency x spacing cells within each quintile of the sequence. These results mirrored the results for the two types of judgments collapsed across positions in the sequence, i.e., continuous judgments of frequency showed no effect of lag at any of the five word positions while each position showed an effect of lag on the terminal test which was similar to that depicted in Figure 3.

Analyses of variance confirmed the appearance of the graphs by indicating that word position did not interact with frequency, spacing, or frequency x spacing for either of the two types of judgments. There was, however, a significant main effect of word position on the terminal judgments with the last word in each cell (i.e., the words appearing in the fifth quintile toward the end of the sequence) receiving higher judgments than the other four words, $\underline{F}(4,92) = 8.85$. There was a similar trend for the continuous judgments but the differences did not attain significance, $\underline{F}(4,92) = 2.28$, $\underline{p} > .05$. The mean judgments of frequency by word position are given in Table 2.

Terminal Test Results vs. Continuous Test Results. A three-way analysis of variance was carried out to compare directly the two measures of retention, continuous judgments of frequency versus terminal judgments of frequency. A summary of this analysis is found in Table C. As expected, there was a significant difference between the two types of

TABLE 2

Mean Judgments of Frequency by Presentation Frequency, Position of Word in Sequence, and Type of Test (Experiment 1)

	Quintile					
Frequency	1	2	3	4	5	$\bar{X}$
			Continuous Test			
2	2.17	2.26	2.38	2.23	2.47	2.30
3	3.21	3.18	3.19	3.13	3.44	3.23
5	4.89	5.22	5.06	5.12	5.09	5.07
$\bar{X}$	3.42	3.55	3.54	3.49	3.66	3.53
			Terminal Test			
2	2.29	2.15	2.26	2.44	2.80	2.39
3	2.91	2.85	2.68	2.69	3.52	2.93
5	3.48	3.56	3.49	4.02	4.08	3.73
$\bar{X}$	2.89	2.85	2.81	3.05	3.47	3.02

frequency judgment, $F(1,23) = 12.42$, $p < .01$. In addition, all interactions involving the testing factor were significant. The frequency x test interaction reflects the fact that the decrease in terminal judgments relative to continuous judgments is especially marked at frequency of 5 while the spacing x test interaction indicates that spacing has a greater effect on terminal judgments than on continuous judgments. The triple interaction arises from the increased effect of spacing over levels of frequency, especially for terminal judgments.

Recognition Measures. Underwood and his colleagues (e.g., Underwood, 1971; Underwood, Zimmerman, & Freund, 1971) have argued that recognition memory is determined to a large extent by a discrimination of situational frequency. In other words, subjects attempt to discriminate those items with situational frequency of one or more from those items with a situation frequency of zero. Thus, "the probability of being correct on a given frequency discrimination corresponds roughly to the probability of being correct on a recognition test under the same conditions" (Underwood et al., 1971, p. 150). If this argument is correct, then the pattern of results for recognition meansures should be similar to the patterns for the judgments of frequency.

Accordingly, the occurrences of recognition misses and false alarms were analyzed. For the continuous judgments, a miss was defined as a judgment of one given on the second occurrence of a word. A false alarm was defined as a judgment exceeding one on the first occurrence of a word. For the terminal judgments, a miss was a judgment of zero given to a test item which in fact was in the sequence and a false alarm was any non-zero judgment given to a "new" word on the test. Mean number of misses and false alarms within various relevant categories are given in Table 3.

For the continuous measures, the overall false alarm rate was 0.14. The number of false alarms in the fifth quintile, i.e., toward the end of the list, exceeded the number in the other four quintiles of the sequency by a significant amoung, $F(4,92) = 4.12$, $p < .01$. This tendency for the false alarm rate to increase toward the end of a sequence has been found in continuous recognition tasks (e.g., Shepard & Teghtsoonian, 1961) and may be related to the increase in judgments (both continuous and terminal) for words in the fifth quintile. The misses on the continuous test occurred at a rate of .04 and showed a tendency to increase (i.e., the hit rate decreased) over spacing, $F(5,115) = 5.05$. This tendency is a replication of previous findings (e.g., Shepard & Teghtsoonian, 1961; Nickerson, 1965) but is opposite to the tendency found here for continuous judgments of frequency to increase over lag.

For the terminal measures, the false alarm rate was .34 but this measure is of comparatively little interest since the number of false alarms here cannot be partitioned

TABLE 3

Derived Measures of Recognition (Experiment 1)

(a) Continuous Test: Mean number of false alarms within each quintile of the sequence (maximum = 18)

Quintile					
1	2	3	4	5	$\overline{X}$
2.17	1.96	2.46	2.33	3.67	2.52

(b) Continuous Test: Mean number of misses by spacing (maximum = 15)

Spacing						
0	1	2	4	8	16	$\overline{X}$
.125	.417	.750	1.13	.542	1.00	.660

(c) Terminal Test: Mean misses by spacing and presentation frequency (maximum = 5)

	Spacing						
Frequency	0	1	2	4	8	16	$\overline{X}$
2	.667	.458	.833	.500	.542	.375	.562
3	.750	.542	.167	.417	.250	.125	.375
5	.667	.208	.333	.250	.250	.083	.299
$\overline{X}$	.694	.403	.444	.389	.347	.194	.412

according to quintiles or frequency x lag cells. The overall miss rate on the terminal test was 0.08. The mean number of misses, shown in Table 3 by frequency x lag cells, indicated a decrease over both spacing and presentation frequency (i.e., the hit rate increased over lag and frequency). For lag, $F(5,115) = 5.18$ and for frequency, $F(2,46) = 7.27$, $p < .01$. The frequency x lag interaction was not significant. The derived terminal recognition measure showed then the same general effect of spacing as the terminal judgments and supports the results of Hintzman et al. (1975b, Experiment 3), as well as the arguments of Underwood et al. (1971).

Conditional Judgments of Frequency. Since lag has a similar effect upon probability of correct recognition and terminal judgments of frequency, the effect of spacing upon terminal judgments may be entirely due to its effect upon recognition. In other words, items appearing at short lags may receive relatively small mean judgments merely because subjects fail to recognize them as often as they recognize items repeated at long lags. This argument was checked by calculating the mean terminal judgments made for those items which were correctly recognized as experimental words. These conditional judgments are depicted in Figure 4. As can be seen, the three curves are very similar to the terminal judgments shown in Figure 3. Hence one can say that spacing has an effect upon judgments of frequency over and above its effect upon probability of correct recognition.

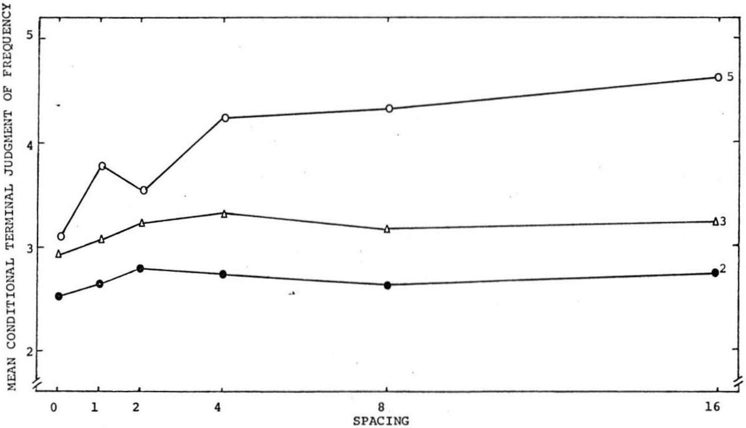

Figure 4. Mean terminal judgment of frequency as a function of spacing and presentation frequency (2, 3, 5), conditional upon correct recognition (Experiment 1).

Discrimination Coefficients. The discrimination coefficient has been defined by Flexser and Bower (1975) as the coefficient of correlation between true and judged frequency. It has been put forward by them as a bias-free measure of subjects' ability to distinguish one frequency from another. Rowe and Rose (1977) reported a massed presentations effect but no other effect of lag when terminal test performance was analyzed in terms of discrimination coefficients. Hence it was decided to test the generality of this finding by using the data from Experiment 1.

Mean discrimination coefficients were found for each level of spacing by calculating for each subect x spacing cell the coefficient of correlation between the true and judged frequency of each word occurring in that cell. The results are shown in Table 4, along with the corresponding results from Rowe and Rose (1977).

The coefficients from Experiment 1 were transformed to Z-scores and an analysis of variance was carried out. A summary of the analysis is contained in Table D. There was a significant effect of lag, $F(5,115) = 8.04$. The comparison of individual means showed that the coeffient at lag of 0 was significantly below the coefficients at all other lags except for lag of 2. The coefficients at lags of 8 and 16 were also significantly greater than the coefficients at lags of 1 and 2. A curvilinear regression analysis showed that the linear and quadratic components of the equation relating

TABLE 4

Mean Discrimination Coefficients

(a) From Experiment 1

Spacing					
0	1	2	4	8	16
.120	.316	.260	.417	.459	.477

(b) From Rowe and Rose (1977)

Spacing			
0	2	16	32
.74	.91	.95	.94

(c) Recalculated from Rowe and Rose (1977)

Spacing			
0	2	16	32
.26	.44	.64	.59

spacing to the discrimination coefficients were both significant, F's(1,115) = 26.26 and 9.55 respectively, p's $<$.01.

Note that the discrimination coefficients are considerably lower here than those found by Rowe and Rose. This discrepancy arises in part from the fact that Rowe and Rose calculated the coefficients from the mean judgments for each frequency x spacing cell whereas here the coefficients were calculated from the judgments of individual words. The method of calculation used here is the same as that used by Flexser and Bower (1975). It is preferable to the procedure used by Rowe and Rose (1977) because judgments of individual words could fluctuate widely around a fairly accurate mean. For this reason, the discrimination coefficients of Rowe and Rose (1977) have been recalculated using the judgments of individual words and appear in Table 4(c). An analysis of variance of these recalculated measures found a significant difference between all pairs (p's $<$.01), except for lag of 16 versus lag of 32.

Discussion

This experiment found that a continuous judgment of frequency test which precedes a terminal judgment test does not eliminate the effect of spacing upon the latter. Rather the lag effect found with the terminal judgments is similar to that found in other task situations. As for the continuous judgments one finds that the spacing had a significant effect only on one frequency x lag cell. This effect was an increase over lag, which is contrary to findings in continuous paradigms measuring recognition. Further evidence from other studies indicates that the minimal lag effect found here may be an anomaly peculiar to this experiment. Following the completion of Experiment 1, Rowe and Rose (1977) had subjects carry out a continuous judgment task followed by a terminal judgment test with presentation frequencies up to seven and lags between repetitions up to 32. They found no overall effect of lag on the final continuous judgments, although there was a tendency for judgments to decline at lag 32 relative to the other values of lag, at least for higher values of frequency. It seems safe to assume then that in general retention does not increase over spacing with continuous judgments as it does for terminal judgments.

Based upon this assumption, one can now say that the typical spacing effect found with the terminal judgment of frequency paradigm arises from some difference between the continuous and terminal tests, since the effect is found with the latter measure but not with the former. One possible factor involved in this difference is the strategy which is adopted in preparation for each type of test. In other words, subjects may use different approaches towards continuous and terminal judgments which could contribute to the effect of spacing. This hypothesis was tested in

Experiment 3. Alternatively, the difference in the effect of lag upon continuous and terminal judgments may involve some fundamental parameter of memory such as retention interval, which is considerably longer for terminal judgments than for continuous judgments. One might argue, then, that terminal judgments are generally lower than continuous judgments because of forgetting over the longer retention interval, especially in the case of items repeated at small spacings. This argument suggests that repetitions at short lags are deficiently processed relative to repetitions at long lags. Hypotheses reflecting this view were tested in Experiments 4 and 5.

First, however, it seemed imperative to gather further evidence of the effect of spacing upon bias-free measures of retention. While Experiment 1 found that non-zero values of spacing affected the bias-free discrimination coefficients, there are still Hintzman's (1969) results to contend with. As mentioned earlier, Hintzman found only a massed presentations effect in a forced-choice test of the frequency of occurrence. Since a forced-choice test is also considered to be free from bias, Hintzman's results are in direct contradiction to the effect of spacing upon discrimination coefficients.

Experiment 2

The purpose of the second experiment was to examine further the massed presentations effect by repeating Experiment 2 of Hintzman (1969) using an amended testing procedure. Hintzman tested words which had appeared twice in his study sequence. Since the effect of spacing in Experiment 1 at non-zero levels of lag was found mainly with those words appearing five times, a forced-choice test using only these words should be more sensitive than the test used by Hintzman.

Subjects in Experiment 2 were given basically the same materials and study procedure as in Experiment 1, i.e., a continuous judgment task. Following this they were given a test consisting of pairs of items and asked to choose the item within each pair which appeared more often in the list. Hintzman (1969) used a similar form of test, except that he arbitrarily designated those words repeated with lags of 4 as standards and those words repeated with lags of 0, 1, 2, 8, or 16 as variables. His test pairs then consisted of one variable and one standard word from the same block of the sequence. Here the critical test pairs contained all possible combinations of items appearing five times but at different levels of lag (i.e., 75 pairs in all). If the effect of spacing beyond a lag of zero was due to some bias factor, then there should be no tendency to choose one item as being more frequent than another, as long as both appeared at non-zero spacings. On the other hand, if the effect of spacing beyond a lag of zero was due to some non-bias factor, then there should be a tendency to choose items appearing at long lags as being more frequent than items appearing at short lags.

Method

Materials. The study items were those used in Experiment 1 with one change. Since only items appearing five times were considered to be crucial, the six packs of cards were constructed such that every pack had the same set of words appearing five times but, within this set, groups of words were shifted such that every word appeared once at each level of lag.

Subjects. The subjects were 24 undergraduates of Memorial University who were paid $3.00 each for their participation. Four subjects were assigned to each of the six packs of cards.

Procedure. The study trial consisted of a continuous judgment task, carried out as in Experiment 1. This task was followed by a test consisting of 100 pairs of items arranged on a sheet in five columns of 20 pairs each. The members of each pair came from the same quintile and the order of the pairs on the test was randomized. The ordinal positions of the short-lag word and the long-lag word within the critical pairs were balanced as closely as possible within each column of the test list. In addition to the 75 critical pairs of words which appeared five times, there were 25 pairs each containing a once-presented word with a word presented 2 or 3 times in the pack. These pairs were included to ensure that subjects were able to discriminate between two different levels of presentation frequency. The ordinal positions of the word types in these pairs were also balanced. There were two versions of the test which were identical except that the ordinal positions of the words in each pair were reversed. The subjects were instructed to underline the word in each pair which in their judgment appeared more often in their pack of cards, guessing if necessary. This final test, like the continuous judgement task, was carried out at the subjects' own pace.

Results

The words presented two or three times were chosen more frequently than the once-presented words on an average of 16.75 (i.e., 67% of the) occasions. Furthermore, 21 of the 24 subjects chose words with a frequency greater than one more often than the once-presented words which, by the sign test, indicates a significant tendency to choose correctly the more frequent word ($p < .001$). Thus the subjects were apparently able to discriminate among levels of presentation frequency.

For the words presented five times, the number of choices made within the pairs of items appearing at each level of lag was counted and subjected to an analysis of variance. This simple computation of the data is justifiable because all of the experimental words appeared equally often at each level of spacing at frequency of 5 and all possible pairings of different levels of lag at frequency of five were tested. The mean number of choices is shown for each level of spacing in Table 5(a). For example, the five words appearing five times in the sequence at a spacing of 2 were chosen as being more frequent than the words appearing five times at each of the other levels of spacing on 12.21 of the 25 occasions. The analysis of variance, which is summarized in Table E(a), indicated that there was a significant effect of spacing in the forced-choice test, ($F(5,115) = 14.21$. An analysis of the individual means showed that all of the words at non-zero spacings were chosen more often than words at lag of 0, $p < .01$. In addition, words at lags of 8 and 16 tended to be preferred to words at lags 1, 2, and 4.

The results from the preceding analysis indicated then that the effect of spacing beyond lag of 0 is not merely due to bias effects. However, it could be argued that the tendency to choose long-lag words over short-lag words was

TABLE 5

Mean Number of Choices Made Within Pairs (Experiment 2)

(a) Including massed items (maximum = 25)

Spacing					
0	1	2	4	8	16
8.46	12.13	12.21	12.67	14.38	15.17

(b) Excluding massed items (maximum = 20)

Spacing				
1	2	4	8	16
9.04	8.92	9.29	11.08	11.67

more apparent than real and was due to a stronger preference for long-lag items over massed items than for short-lag items over massed items. In order to check this possibility, the choices within pairs were counted with item appearing at zero spacings eliminated. The mean number of choices are shown in Table 5(b) and the summary of the analysis of these results is given in Table E(b). Once again there was a significant effect of spacing, $F(4,92) = 5.40$, with the analysis of individual means indicating that words at lags of 8 and 16 were chosen in preference to words appearing at lags 1, 2, and 4. There were no significant preferences with these two spacing sub-groups.

Discussion

The data found in this experiment showed that spacing had a significant effect beyond a mere massed presentations effect when a bias-free forced-choice test was used. As such, they replicated the results of Experiment 1 where the discrimination coefficients were used and led to the conclusion that the tendency for judgments of frequency to increase over spacing is not due to some bias effect.

The results here did not replicate those of Hintzman (1969, Experiment 2). This failure to support Hintzman probably reflects a lack of sensitivity in his test. He used a subset of all possible pairings across spacing and his test words appeared no more than twice in the sequence. The experiment here provided a more sensitive test of the effect of bias factors in that all possible pairings across lag of words appearing five times were used.

In conclusion then, a dichotomous view of the spacing effect was not supported. Experiment 2 did not of course preclude the possibility that other evidence may indicate a massed presentations effect. Such an effect, however would not appear to indicate a "true" memory deficit for massed repetitions and response bias at non-zero lags. Therefore the remainder of this thesis accepted the view that spacing affects massed and distributed repetitions in basically similar ways, although to different degrees. Experiment 3 was carried out to examine the hypothesis that test strategy contributes to the effect of spacing.

Experiment 3

Comparison of the final judgments of frequency on the continuous test and the terminal judgments of frequency in Experiment 1 leads one to infer that the spacing effect arises from some difference between the two types of test. Although the two tests require subjects to carry out essentially the same task, namely to make judgments of frequency, the continuous judgments constitute an ongoing task whereas the terminal judgments are made after the subject has been through the entire sequence. This judgment from a closed set may precipitate a strategy on the terminal test which differs from the strategy used on the continuous test.

Experiment 3 examined the test strategy hypothesis by employing a group of experimental subjects for whom the terminal test items appeared singly on index cards which were added on to the study sequence without any additional instructions. Thus the experimental group appeared to be engaged in one long continuous judgment test. The instructions used in Experiment 3 required subjects to judge the number of previous occurrences of each item within the list. A second group of subjects, who essentially repeated the procedure of Experiment 1 but with the new instructions, was employed for comparison purposes. The results of the experimental group were compared then to the results of the comparison group to determine whether the effect of spacing was similar in the two conditions.

If the two groups showed different effects of spacing on the terminal test items, one would conclude that the subjects' approach to the terminal task differed from their approach to the continuous task, such that the effect of spacing was more profound with the former. In this case, further investigation of these subjective approaches would have to be carried out. On the other hand, if the two groups showed similar results on the terminal test items, then one would conclude that the spacing effect and, incidentally the general decline in terminal judgments relative to continuous judgments were due to the retention interval between an item's last appearance on the continuous test and its occurrence on the terminal test. In this experiment the average retention interval was about 145 items while the maximum spacing between repetitions on the continuous test was only 32.

To summarize then, Experiment 3 was carried out to test whether giving subjects a terminal judgment of frequency test which was clearly separated from the continuous judgment test would interact with the factor of spacing. At the moment there are reasons to support each of two expectations: (1) both comparison subjects and experimental subjects would show the typical terminal judgment functions, or (2) the experimental subjects would show no significant effect of spacing on the terminal test.

Method

Materials. The list used in the continuous task was the 235 list employed by Rowe and Rose (1977). This list consisted of words chosen from the same population as those words used in Experiment 1. As before, the words appeared 2, 3, or 5 times but this time the spacings had values of 0, 2, 16, or 32 and there were only four items at each frequency x spacing cell distributed equally throughout the sequence. In those cases of higher levels of frequency and long lags, this even distribution was not possible but was approximated.

Eight packs of cards were made such that the sets of four words allotted to the frequency x spacing cells were rotated from cell to cell across the packs. This was done in such a way that each repeated word appeared twice at each level of spacing and in approximately equal proportions at each level of frequency.

There were 80 test items which consisted of all 48 words which were repeated in the continuous test sequence, 16 randomly chosen filler items, and 16 "new" words which did not appear in the sequence but were chosen from the same population. For the comparison group, the test items appeared as before on a sheet, arranged in four columns of 20 items each, such that an equal number of each of the three types of test items appeared in each half of the test. For the experimental group, these test items were typed singly onto index cards and appended onto the main study sequence. The order of the test items for the experimental group was the same as the order for the comparison group going down the columns from left to right on the test sheet.

The materials used by Rowe and Rose (1977) were chosen because it was considered that the continuous task sequence would appear to be too long for the experimental group (327 items + 80 test items) if the list from Experiment 1 were used. The 235 list from Rowe and Rose contained only 211 items which produced a sequence of seemingly manageable length after the 80 test items were appended.

Subjects. The subjects were 64 undergraduates of Memorial University who were paid $3.00 each for their participation. The subjects were assigned randomly in equal numbers to the two groups. Four subjects were assigned to each of the eight packs of cards. The testing was carried out in small groups of not more than six subjects each.

Procedure. As mentioned above, the procedure for the continuous judgment task was the same as in Experiment 1, except that subjects were asked to judge the number of previous occurrences of each word within the sequence. In other words, the first time a subject met a word he wrote "0" on the response sheet, the second time he met the same word he wrote "1", etc. The subjects were asked to rate the number of previous occurrences of an item on the continuous test in order that the judgments made by the two groups on the 80 terminal test items would be comparable. If the subjects included within their ratings the current occurrence, as in Experiment 1, then the ratings of the experimental group would be expected to be higher than those of the comparison group. As before, the subjects had an initial 12 practice words and half of them went through the main sequence in reverse order.

The task for the experimental group consisted solely of making continuous judgments of frequency for the 291 items in their pack of cards. The comparison group made continuous judgments of the 211 items in their pack followed by terminal judgments of the 80 test items, made under conditions identical to Experiment 1.

Results

Continuous Judgments. The mean final judgments on the continuous test are given for each group in Table 6 and are shown in Figure 5. A summary of the analysis of variance is given in Table F. As expected, there were no significant differences between the group doing a long continuous task and the group given a shorter continuous task followed by a terminal test, nor were any interactions involving the factor of group significant. Significant effects were produced only by presentation frequency, $F(2,124) = 837.06$, and the frequency x spacing interaction, $F(6,372) = 8.83$. This significant interaction seems to reflect the tendency for the continuous judgments to increase over spacing at frequency of 2 while it does the opposite at frequency of 5.

TABLE 6

Mean Final Continuous Judgments of Frequency by Group (Experiment 3)

	Spacing				
Frequency	0	2	16	32	$\overline{X}$
	Experimental Group				
2	1.13	1.27	1.55	1.58	1.38
3	2.24	2.33	2.56	2.18	2.33
5	4.06	3.81	3.90	3.85	3.91
$\overline{X}$	2.48	2.47	2.67	2.54	2.54
	Comparison Group				
2	1.23	1.32	1.37	1.39	1.33
3	2.24	2.26	2.44	2.18	2.28
5	4.13	4.02	3.85	3.64	3.91
$\overline{X}$	2.53	2.53	2.55	2.40	2.50

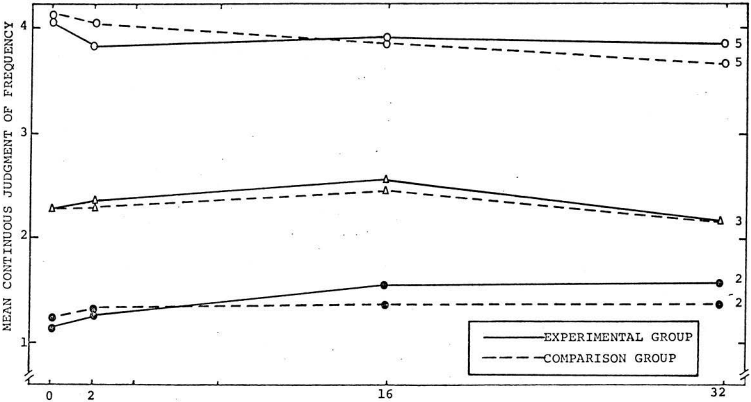

Figure 5. Mean final continuous judgment of frequency as a function of spacing, presentation frequency (2, 3, 5), and group (Experiment 3).

Terminal Judgments. The mean terminal judgment for group at each spacing x frequency cell is shown in Table 7 and Figure 6. A summary of the analysis of variance is shown in Table G. As usual, significant effects were produced by frequency, $F(2,124) = 234.88$, by lag, $F(3,186) = 66.61$, and by the frequency x lag interaction, $F(6,372) = 10.98$. The results also showed that the type of group did not have a significant effect, $F(1,62) = 2.15$, nor, more importantly, did any interaction involving the group factor.[1] Hence the effect of spacing was the same upon the two groups.

Discussion

The results from the final continuous judgments showed no significant difference between the experiment and comparison groups, hence confirming that the groups could be legitimately compared on terminal judgments. They also confirmed the assumption from Experiment 1 that retention does not generally increase over spacing with continuous judgments as it does for terminal judgments. In fact, the continuous judgments in Experiment 3 showed a small general tendency to be undestimates at a frequency of 5. Hence the significant effect of lag at frequency level 5 in Experiment 1 seems indeed to be artifactual.

---

[1] Because the group factor had neither a major nor an interactive effect, regression analyses were not carried out since these would add little to those carried out in Experiment 1.

TABLE 7

Mean Terminal Judgments of Frequency by Group
(Experiment 3)

	Spacing				
Frequency	0	2	16	32	$\overline{X}$
		Experimental Group			
2	1.66	1.91	2.12	2.18	1.97
3	1.88	2.45	2.68	2.55	2.39
5	2.79	3.20	3.88	4.09	3.49
$\overline{X}$	2.11	2.52	2.89	2.94	2.62
		Comparison Group			
2	1.38	1.76	1.77	1.69	1.65
3	1.70	2.27	2.39	2.35	2.18
5	2.06	3.35	3.51	3.92	3.21
$\overline{X}$	1.71	2.46	2.56	2.65	2.35

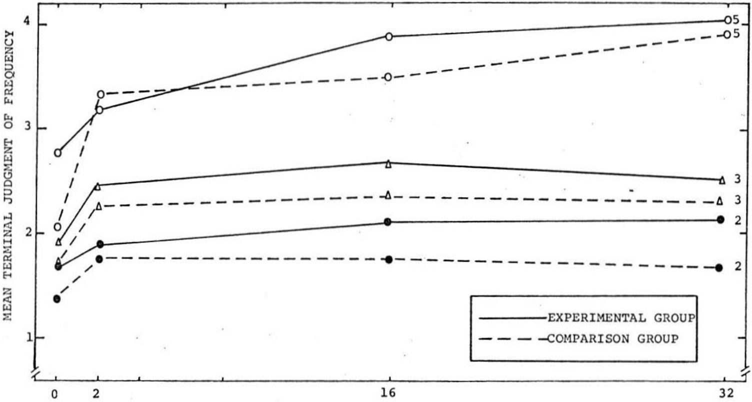

Figure 6. Mean terminal judgment of frequency as a function of spacing, presentation frequency (2, 3, 5) and group (Experiment 3).

Turning to the terminal judgments of frequency, one finds for both groups the usual effects of spacing and presented frequency. These results indicate then that the spacing effect does not arise because of some strategy which subjects adopt when faced with a terminal test. Rather, the results suggest that the spacing effect (and incidentally the tendency to underestimate the frequency as presentation frequency increases) arises from the task of making judgments of frequency after a comparatively long retention interval. Glenberg (1976) has shown that the length of the retention interval affects the form of the spacing function. The results of Experiment 3 strengthen this position by indicating that the length of the retention interval is the <u>only</u> major difference between the continuous and terminal judgments of frequency. In particular, the "strength" of the items which were repeated originally at short lags has a relatively low value after a long retention interval and hence those items are relatively underestimated. This in turn suggests, as was mentioned in the introduction to the research, that items repeated at low values of spacing receive inferior processing relative to those repeated at high values of spacing. This suggestion will be examined in Experiment 4.

Experiment 4

The results of the first three experiments of this thesis support the general view that the spacing effect is due to deficient processing of items repeated at short lags and not due to some bias effect or control process. This view in turn suggests, as has been mentioned before, that processing short-lag items to a "deep" level should obviate any deficiency in processing. However, merely giving subjects an incidental semantic task on the study trial will not elevate the probability of retention of short-lag items relative to long-lag items and eliminate the spacing effect. Rather, the retention function as a whole is elevated, relative to the retention function following a non-semantic task, and the spacing effect remains (Rose & Rowe, 1976).

The reason for this seems obvious. When subjects are engaged in a semantic rating task such as the rating of implied strength or goodness used by Rose and Rowe (1976), they need not process repeated items to the same level as they process items on their initial presentations, at least not when the repetitions occur at short spacings. In such cases the subjects can probably recall their initial ratings and merely report these memories, rather than repeat the whole rating process. This argument is analogous to the view of Lockhart (Note 1) that subjects merely scan recent episodic memory when repetitions are close together but reconstruct something approaching the encoding of the original event as spacings increase.

One plausible solution to this problem is to require a different rating task for each presentation of the same item. This procedure should remove the possibility of subjects recalling their initial rating instead of carrying out the process required to achieve that rating. Shaughnessy (1976) carried out this type of test and found that asking subjects to perform different rating tasks (frequency in printed English; imageability; connotative pleasantness) on the two or three presentations of a word did not eliminate the spacing effect on a subsequent free recall test.

However, there are methodological characteristics of Shaughnessy's study which may have resulted in an inadequate test of the levels of processing hypothesis. As he himself points out, subjects making semantic ratings do not do so in isolation but rather compare the current item to other items. In particular, he suggests that these comparison items are likely to be other items in the sequence, especially items repeated at long lags, if for no other reason than that long-lag items are spread more throughout the sequence and are hence available as references to a greater number of other items than are short-lag items. As a result, long-lag items may develop more associative links, receive more processing, etc. The upshot is that a spacing effect remains.

Also, it is possible that previous ratings of an item may influence the current rating of that item, even if these ratings are along different dimensions. For instance, Paivio (1975) found that the probability of recall of twice-presented consecutive words, which had been rated for imageability on one presentation and rated for pleasantness on the other presentation, was less than expected from the recall of once-presented words rated for either imageability or pleasantness. This result was interpreted as indicating that these two rating tasks involved processes which were not mutually independent. If subjects do consult their previous rating of an item when they are rating a repetition of that item along a different dimension, then we have a situation comparable to a task requiring subjects to rate all presentations of an item along the same dimension. In other words, the subjects can once more scan their recent memory when the items are repeated at short lags but must use the deeper process when the items are repeated at long lags.

An alternative to Shaughnessy's solution involves the use of an orienting task in which each presentation of an item is presented in as much isolation as possible from all other presentations of both that item and other items. In this way, each occurrence of an item should be processed to the same depth and associations should not develop with long-lag items in preference to short-lag items. In Experiment 4 subjects on the study trial were asked binary questions of a semantic nature concerning the items in the sequence. Such questions, requiring as answers "yes" or "no", have been used before in incidental learning paradigms (see Craik & Tulving, 1975 for example). Furthermore, in one condition, a different question was asked for each item in the sequence, whether it was a repetition or not. In this way, non-random preferential associations among items within the sequence or among repetitions of items should not develop. In addition, the questions asked of repetitions under the different-question condition, biassed the same meaning of a repeated word. For example the questions accompanying the word "earth" all concerned "earth" as a planet, not "earth" as soil, etc. In this way, the verbal contexts were considered to differ under the different-question condition mainly in a non-semantic sense.

In a second condition, the same binary question was asked for each presentation of the same item but different questions were asked of different items. This procedure was intended to reduce the tendency to form associations among different items but not force subjects to process the various presentations of the same item differentially.

Experiment 4 was designed essentially to test two theoretical explanations of the spacing effect: the non-semantic form of the encoding variability hypothesis and the levels of processing hypothesis. It has already been argued that most evidence goes against the encoding variability hypothesis if quite separate meanings of a repeated word are biassed by the encoding contexts. However, non-semantic versions of the encoding variability hypothesis continue to flourish (Glenberg, 1976).

According to the encoding variability hypothesis, the "typical" spacing effect as depicted by the curve AB in Figure 7 occurs because the contexts of the presentations of an item repeated at long lags are more varied than the contexts of an item repeated at short lags. Hence the greater the variability of the encoding contexts, the greater the retention. Two predictions follow directly from this view. One is that forcing subjects to encode each repetition of an item in a different context should eliminate the spacing effect and produce a high level of retention as shown by the horizontal line through B in Figure 7. The second prediction is the converse of the first, namely that forcing subjects to encode every presentation of a repeated item in the same context should also eliminate the spacing effect but at a low level of retention as depicted by the horizontal line through A.

Strictly speaking of course, one cannot have complete experimental control over encoding contexts. If temporal factors and neighbouring items in the list are important components of the encoding context, then a

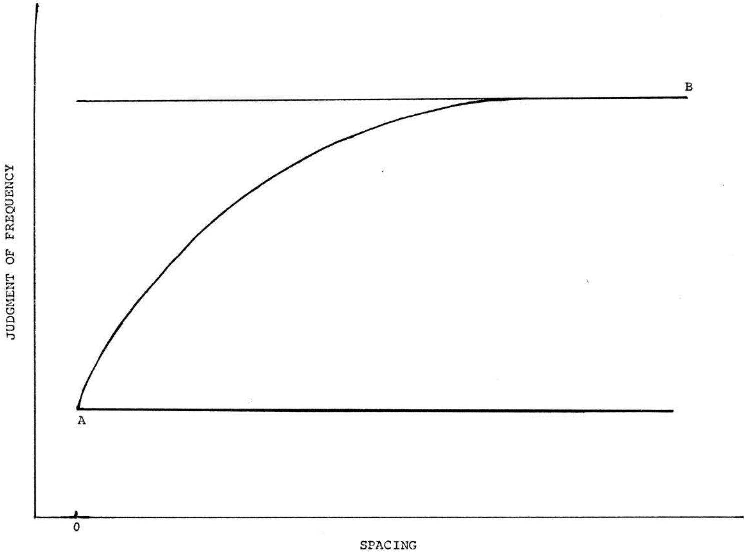

Figure 7. Hypothetical spacing functions predicted in Experiment 4 (see text for explanation).

spacing effect could still occur in spite of the preceding contextual manipulations. For instance, temporal differences are obviously greater among items repeated at long lags than among items repeated at short lags. However, one would predict from the encoding variability hypothesis that forcing the same encoding of each presentation of an item would at least decrease the effect of spacing relative to a free, uncontrolled situation.

Turning to the levels of processing hypothesis, the spacing effect is said to occur because items repeated at relatively long lags have to be re-processed to the original "deep" level, whereas items repeated at short lags are encoded by merely scanning recent short-term memory. Hence the necessary condition for good retention is not variable encoding but encoding to a deep level. The levels of processing hypothesis predicts along with the variable encoding hypothesis that the different-question condition of Experiment 4 will produce results approximating the straight line through B in Figure 7. This outcome should occur however because subjects under this condition will not be able to scan recent memory for an answer but will have to process each question separately, regardless of the spacing between repetitions.

The difference between the two hypotheses lies in the prediction of the outcome for the same-question condition. The encoding variability hypothesis predicts that encoding context on all presentations of a word should attenuate the effect of spacing and produce a relatively low level of retention as shown by the straight line through A. The levels of processing hypothesis, on the other hand, leads one to expect the usual spacing function as depicted by the curve AB. This outcome follows from shallow scanning of recent memory when answering a question repeated after a short lag and a deep reconstruction of the original cognitive process when answering a question repeated after a long lag.

To summarize the foregoing arguments with respect to an analysis of variance involving factors of question condition, presentation frequency, and spacing, the encoding variability hypothesis predicts that different-question condition will exceed the same-question condition considerably, that there will be little or no overall effect of spacing, and that no significant question condition x spacing interaction will emerge. The levels of processing hypothesis predicts that the different-question condition will only moderately exceed the same-question condition overall, that there will be a moderate effect of spacing (due to the same-question condition mainly), and that there will be a strong question condition x spacing interaction.

Method

Materials. The study words were chosen from the same population as those words used in the previous experiments, except that one seven-lettered word (college) was inadvertently included. However, the sample used here contained several changes from the previous ones because each experimental word had to produce six meaningful questions, three requiring positive answers and three negative, all of which biassed the same meaning of the word. Examples of the types of questions used in Experiments 4 and 5 are found in Table 8.

As before, each subjects was given a pack of cards, each of which contained a word typed in capital letters with a simple binary question pertaining to that word typed in lower case beneath it. The construction of the pack paralleled that of Experiment 1, i.e., words appeared 2, 3, or 5 times at spacings of 0, 1, 2, 4, 8, or 16 with 27 filler items appearing once each. Four words were assigned to each frequency x spacing cell, which produced 267 cards per pack. This change from Experiment 1 allowed an equal application of the two within-subjects experimental conditions (i.e., two words per condition) and compensated for the fact that the task in Experiment 4 required more time per word than the task in Experiment 1. According to the analyses of the judgments of frequency by word position carried out on the data from Experiment 1, this allotment of only four words to each frequency x spacing cell should not alter the spacing function.

TABLE 8

Examples of Questions Used in Experiments 4 and 5

(a) Experiment 4 (Semantic Questions)		
	1. BUTTER	Is this a dairy product?
	2. DEATH	Does this often result from a headache?
	3. MARKET	Is this a place for buying things?
(b) Experiment 5 (Orthographic/Graphemic Questions)		
	1. COTTON	Does this word contain four different letters?
	2. board	Is this word typed in upper case letters?
	3. MONTH	Is this word written in black ink?

Type of question condition was applied as a within-subject variable. Within each pack, half of the experimental words were accompanied by different question on each presentation of the word while half were accompanied by identical questions on each presentation of the same word. As was mentioned before, each word had its own unique set of questions and care was also taken not to repeat an experimental word in the question of another word. Within the different-question condition, half of the questions were positive and half were negative. For the two words appearing three times each in a frequency x spacing cell in the different-question condition, one word had one negative and two positive questions and the second word had one positive and two negative questions. For the two words appearing 5 times in a cell under the different-question condition, the split was 3:2 and 2:3. In these conditions, the positive and negative questions occupied alternate ordinal positions. For the two words appearing in each frequency x spacing cell under the same-question condition, one word was accompanied on all presentations by the same positive question while the second word was always accompanied by the same negative question. Filler words of course could not be assigned to one condition or another. Rather, half of them were accompanied by question requiring a positive answer while the other half required a negative answer.

As in Experiment 1, there were six packs of cards constructed in order to counterbalance specific item effects across frequencies and spacings. The question conditions and the levels within these conditions were also allotted in a counterbalanced fashion across the packs.

Subjects. The subjects were 48 undergraduates of Memorial University who were paid $3.00 each for their participation. Eight subjects were assigned to each of the six packs of cards. The testing was carried out in small groups of not more than six subjects each.

Procedure. The subjects were given an instruction sheet which informed them that the experiment was concerned with people's view of a number of common concepts. Accordingly, they would be required to answer with "yes" or "no" a number of simple questions relating to these concepts. Following 12 words of practice, the subjects went through the pack of cards at their own pace and wrote the answer to each question in the appropriate space on a response sheet. Half of the subjects went through their pack in reverse order.

When they had finished the question task, the subjects were given an unexpected test requiring them to judge how often each of the words on the test sheet had appeared in the pack of cards. The test consisted of all 72 experimental words plus 14 new words chosen from the experimental population and 14 filler items. The fillers were randomly chosen from those once-presented words which did not occupy primacy or recency positions in the pack.

the judgments of frequency test was unpaced. The entire experiment required about 40 minutes.

Results

Judgments of Frequency: Main Results. The mean judgment of frequency for each frequency x spacing cell within each condition is shown in Table 9 and depicted in Figures 8 and 9. The initial analysis of variance involved the factors of presentations frequency, spacing, and question condition and is summarized in Table H. The analysis showed the usual strong effect of frequency, $F(2,94) = 173.29$, and also a significant effect of spacing, $F(5,235) = 21.27$. There was no effect of question condition, ($F < 1$) but there were significant interactions involving the factor of question condition. The two important interactions were question condition x spacing, $F(5,235) = 8.42$, which reflected the greater effect of spacing under the same-question condition relative to the different question condition and the triple interaction, $F(10,470) = 2.15$, $p < .025$.

The significant triple interaction indicated different frequency x spacing patterns for the two question conditions and justified separate analyses of the judgments at each level of frequency. Significant differences between the two question conditions were found at frequencies of 3 and 5, F's$(1,94) = 4.42$ and 6.04 respectively, $p < .05$. In addition, there were significant question condition x spacing

TABLE 9

Mean Judgments of Frequency by Question Condition
(Experiment 4)

	Spacing						
Frequency	0	1	2	4	8	16	$\bar{X}$
			Different Questions				
2	2.53	2.21	2.26	2.31	2.35	2.61	2.38
3	2.98	3.30	3.65	3.47	3.76	3.30	3.41
5	4.04	4.06	4.62	4.27	4.90	4.67	4.43
$\bar{X}$	3.18	3.19	3.51	3.35	3.67	3.53	3.40
			Same Questions				
2	1.61	2.70	2.54	2.77	2.67	3.19	2.58
3	2.42	2.63	3.03	3.36	3.71	3.98	3.19
5	3.78	3.97	4.48	4.35	5.72	6.09	4.73
$\bar{X}$	2.60	3.10	3.35	3.49	4.03	4.42	3.50

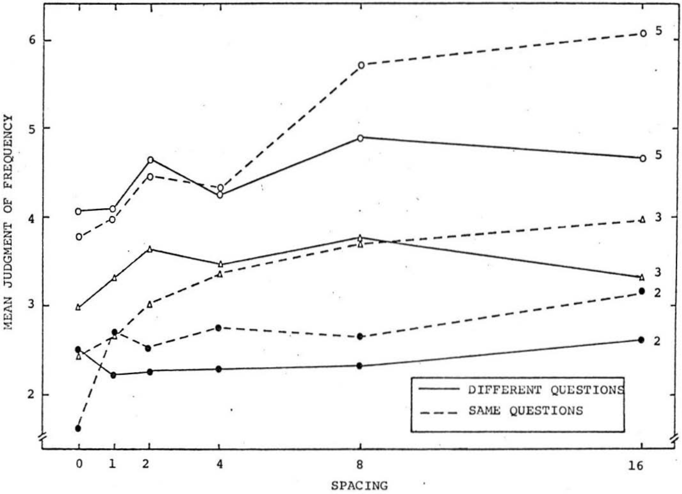

Figure 8. Mean judgment of frequency as a function of question condition, presentation frequency (2, 3, 5), and spacing (Experiment 4).

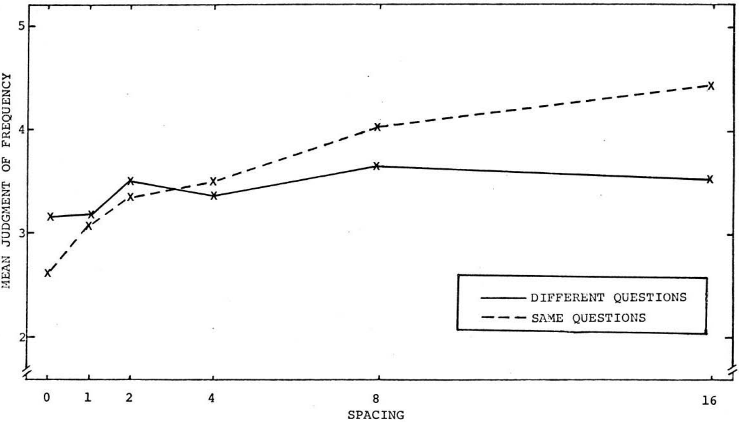

Figure 9. Mean judgment of frequency by question condition and spacing, collapsed across presentation frequency (Experiment 4).

interactions at all levels of presentation frequency, $\underline{F}$'s(5,470) > 4.87, $\underline{p}$'s < .01.

Because of these significant interactions, further analyses were run at each level of frequency for each question condition separately. For the different-question condition, there was no significant effect of spacing at frequency of 2, but there was at frequencies of 3 and 5, $\underline{F}$'s(5,470) > 3.02, $\underline{p}$'s < .025. A comparison of the individual means indicated that, for frequency of 3 the mean judgments at lags 2 and 8 exceeded that at lag of 0 while, for frequency of 5, the mean judgment at lag 8 exceeded that at lag of 0 with borderline significance. Turning to the same-question condition, the analyses showed a significant effect of spacing at each level of frequency, $\underline{F}$'s(5,470) > 8.47. The comparison of individual means showed only a massed presentations effect at frequency of 2, $\underline{p}$ < .01), while at frequency level 5, the effect of spacing was at lags 8 and 16 while the mean judgments exceeded all others ($\underline{p}$ < .01) but did not differ between themselves. At frequency of 3, no level of spacing differed from the adjacent level but all other pairs of comparisons were significantly different.

Curvilinear regression analyses were also carried out for each condition at each level of frequency. At frequency of 2, the different-question condition showed no regression of judgments of frequency upon spacing while the same-question condition showed significant linear and cubic components, F's(1,470) = 21.22 and 10.55 respectively. At frequency of 3 the different-question condition showed only a significant quadratic component, $F(1,470) = 9.94$, $p < .01$, while the same-question condition showed significant linear and quadratic components, F's(1,470) = 48.70 and 8.99 respectively, p's $< .01$. At frequency of 5, the linear and quadratic components were significant under both question conditions, $F(1,470) = 8.52$ and 5.19 respectively, p's $< .025$, for the different-question condition and F's(1,470) = 126.40 and 8.53 respectively, p's $< .01$, for the same-question condition.

In summary then, the judgments of frequency found here did not support the encoding variability hypothesis. The unexpectedly high mean judgments for the same-question condition, the overall spacing effect, and the question condition x spacing interactions all contradicted predictions derived from it. Alternatively, the question condition x spacing interactions and the attenuated effect of spacing under the different-question condition supported the levels of processing hypothesis, although the mean judgments under the same-question condition were generally higher than expected.

Judgments of Frequency: Positive vs. Negative Questions. Craik and Tulving have shown that long-term recall is often lower when negative questions are asked in an incidental learning task than when positive questions are asked. Therefore, the judgments of frequency were analyzed for the positive and negative question separately. As far as the different-question condition was concerned, there was no difference in mean judgments at frequency of 3 between two positive questions with one negative question and two negative questions with one positive question. The means for these two sub-conditions were 3.41 and 3.42 and the factor of spacing showed a similar trend in each case. The two sub-conditions at frequency of 5 also showed no differential effect of spacing, although the means differed slightly at 4.38 and 4.47 for the three positive questions and three negative questions respectively.

The cell means for the same-question condition are shown in Figure 10 for positive questions and negative questions separately. An analysis of variance showed that the mean judgment of 3.54 for the positive question sub-condition was not significantly larger than the mean judgment of 3.46 for the negative question sub-condition, $F(1,47) < 1$. In addition, the effect of spacing was similar in each case, as indicated by the non-significant interaction involving the factor of question type, F's < 1.81 in all cases. Hence the type of question asked had no confounding effect for the purposes of this experiment.

Judgments of Frequency for Words at Frequency of 1 vs. Words at Lag of 0. Craik and Lockhart (1972) state that repetition of analyses which have already been carried out will not enhance memory. This implies that repetitions at short spacings of a word accompanied by the same question

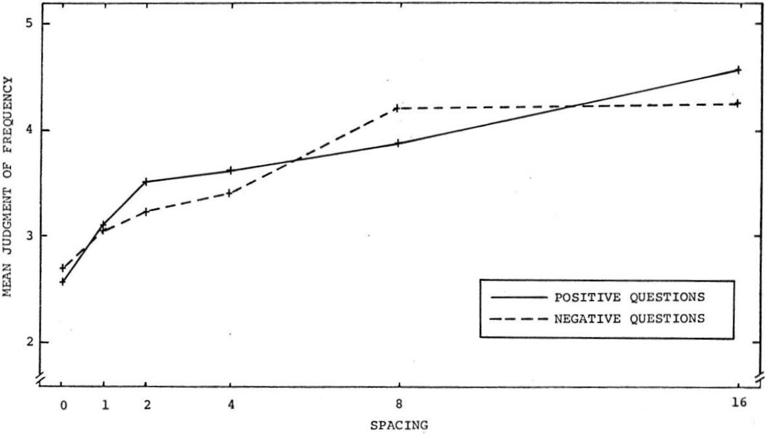

Figure 10. Mean judgments of frequency for condition S by type of question and spacing, collapsed across presentation frequency (Experiment 4).

would add nothing significant to a single presentation of that word. Glenberg (1976) also states that a repetition of an item at very short lags will produce no learning beyond that produced by a single presentation. This view is represented in Figure 1 of this thesis by a single circle to represent two massed presentations of an item. It was therefore considered informative to compare the judgments of frequency of words presented once with judgments of words presented two or more times at zero spacings under the same-question condition.

The mean judgments at lag of 0 for the same-question condition can be seen in Table 9. The mean judgments for once-presented words was 1.54. An analysis of variance showed a significant difference among these means, $F(3,141) = 46.06$. The mean for once-presented words did not differ from that of massed words at frequency of 2 but all other pairs were significantly different ($p < .01$). These results then go against Craik and Lockhart (1972) and the assumption of Glenberg (1976) but agree with Nelson (1977), who found that two massed presentations led to better recall than a single presentation.

Recognition Scores. Rowe (1973a, b) found that judgments of frequency of words repeated up to five times in exactly the same phrase tended to be higher than the judgments of words repeated in "similar" phrases, i.e., phrases which were different but biassed the same meaning of the experimental word. The judgments of frequency for

Experiment 4 showed that the same-question condition exceeded the different-question condition at higher levels of spacing, which is entirely consistent with Rowe who used random lags with a mean value of 25 (1973a) or 34 (1973b). In the same experiments, Rowe also found that the verbal context of the experimental words had no effect upon their recognition. In order to compare this finding with the present results, derived recognition scores were calculated by considering that any experimental word given a judgment of zero was a "miss" and all other judgments of an experimental word were "hits".

The probability of correct recognition is shown by conditions collapsed across frequency in Figure 11. It can be seen that recognition was consistently higher for the words in the different-question condition than in the same-question condition. An analysis of variance confirmed that question type had a significant effect, $\underline{F}(1,47) = 19.81$. There were no other significant effects except for presentation frequency, $\underline{F}(2,94) = 5.81$, $\underline{p} < .01$. The recognition results here then disagree with those of Rowe (1973a, b). However, the greatest differences lie at small values of lag and it is possible that no significant differences would exist if the long lags used by Rowe were also used here. More interestingly, the pattern of results found here (and in Rowe, 1973a, b) for the recognition hits differed from the pattern for the judgments of frequency, where the same-question condition generally exceeded the different-question

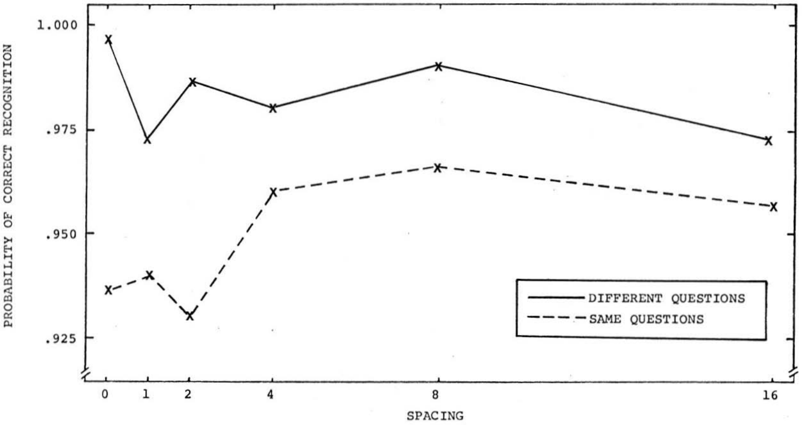

Figure 11. Mean probability of correct recognition by question condition and spacing collapsed across presentation frequency (Experiment 4).

condition at lags of 4 or more. As such then, the arguments of Underwood and his colleagues that probability of correct recognition should correspond with judgments of frequency when tested under the same conditions were not upheld. More will be made of this point in the discussion.

The recognition results for the same-question condition were also examined for positive vs. negative questions. Unlike the judgments of frequency, there was a difference here with the negative questions producing 66 misses (hit rate of .92) compared to only 27 misses (hit rate of .97) for the positive questions, $F(1,47) = 18.38$. The weak tendency for a spacing effect in the same-question condition was also due mainly to the results of the negative questions. However, since no other factor or interaction had a significant effect, the effect of type of question upon recognition was not crucial.

The probability of recognition of once-presented words was compared to the probabilities of recognition of words repeated at zero spacings. These probabilities were nearly identical at 0.932 for words presented once and 0.937 for words repeated at zero lags (collapsed across frequency). There were no significant differences among the four levels of frequency, $F(3,141) = 1.47$. As such these results differ from the judgments of frequency for once-presented and massed words.

Finally, it is realized that hit rates should not be considered in isolation from the probability of making a false alarm, which in Experiment 4 was 0.20. Unfortunately, the nature of the paradigm used here precludes the calculation of separate false alarm rates for the two experimental condition. However, a pilot study which preceded this experiment used type of question condition as a between-subjects variable and found similar false alarm rates of 0.16 and 0.13 for different-question and same-question conditions respectively ($t_{38} = 0.57$, $p < .025$). Hence one is probably safe in assuming that the hit rates found here are free from the effects of differential response bias.

Conditional Judgments of Frequency. Judgments of frequency conditional upon correct recognition were calculated for each question condition x frequency x spacing cell. These yielded patterns which were very similar to those depicted in Figure 8, except that the judgments for the same-question condition tended to be slightly higher relative to the judgments for the different-question condition. These tendencies can be seen clearly by comparing Figure 12, which depicts the conditional judgments collapsed across frequency, with Figure 9. This comparison shows that the different-question condition exceeds the same-question condition on "pure" judgment of frequency only at zero spacings.

Incidental Tasks. The designation of questions as positive or negative in this experiment was based upon the judgment of the experimenter. The subject might of course judge otherwise (or might answer erroneously). Accordingly,

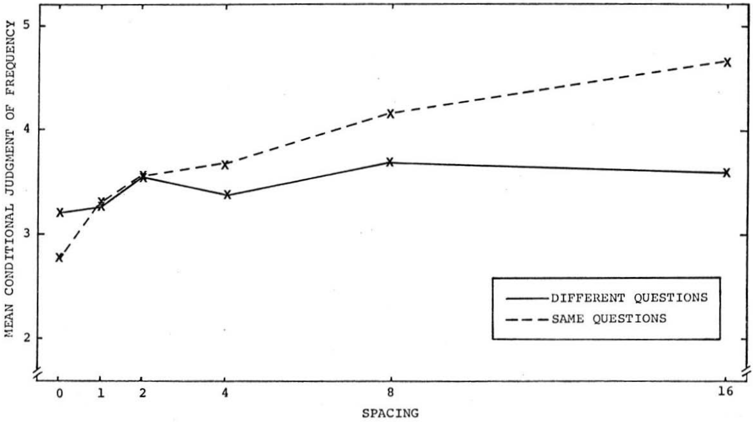

Figure 12. Mean judgment of frequency by question condition and spacing, conditional upon correct recognition (Experiment 4).

error rates on the incidental tasks in the pilot study were examined and the judgments by question type (i.e., positive vs. negative) were adjusted according to the answers given by the subjects. The errors rates were 4.67 percent for the different-question condition and 5.06 for the same-question condition. The adjustments to the judgments made little difference to the cell means, i.e., the positive and negative errors tended to cancel out. For this reason, the incidental tasks were not scored in Experiment 4. Rather they were inspected to ensure that the subjects responded to all questions and to ensure that they performed the task seriously, i.e., that they did not write down inappropriately long sequences of "yes" or "no" or alternations between these two answers. No subject appeared to respond inappropriately.

Discussion

The judgments of frequency found in this experiment support the levels of processing hypothesis in so far as the question condition x spacing interactions were significant and the spacing effect under the different-question condition was attenuated. The different-question condition then is considered to have induced moderately deep processing at all levels of spacing by precluding the use of the shallow scanning process. On the other hand, the same-question condition induced processing at a level which became progressively deeper over lag and thereby produced the spacing effect.

The different-question condition was predicted to exceed the same-question condition at short lags and to equal it at some asymptotic value at long lags. This prediction was supported when retention was measured by probability of correct recognition but, unexpectedly, the judgments under the same-question condition equalled or exceeded the judgments under the different-question condition at all spacings except zero. The different patterns produced by these two measures would appear to reflect the different demands made by recognition and judgments of frequency. Recognition of an item presented n times on a learning trial requires the retrieval on a test of only <u>one</u> of the n traces of that item. The judgment of frequency of the same item requires the retrieval of <u>all</u> n traces.

As a reconciliation of these differences, it is suggested that the test items in Experiment 4 are processed as follows. When a subject encounters an item on the judgment of frequency test, an attempt is made to reconstruct the encoding of that item as it occurred on the learning trial. An item occurring under the different-question condition is processed moderately deeply, regardless of lag, and its trace is therefore quite durable. Hence it is relatively easy for the reconstructive process to make contact with at least <u>one</u> of these durable traces and therefore recognition of items learned under the different-question condition is good. At the same time, however, an item presented n times under the different-question condition has n distinctive traces, and the retrieval of <u>all</u> of these distinctive traces is relatively difficult because they share fewer common attributes than traces under the same-question condition. As a result, judgments of frequency are relatively low when different questions are asked, especially at higher levels of spacing.

Consider now the situation when the test items was encoded n times with the same question. At short lags, there is one deeply encoded trace (corresponding to the initial presentation) and n-1 poorly encoded traces. Here recognition is poor because there are fewer "strong" traces with which to make contact than under the different questions condition. Although the n traces under the same-question condition are strongly linked via many common attributes, frequency judgments are low at short lags because the bonds are among relatively weak traces. However, at long lags, items studied under the same-question condition lay down closely linked traces which are now deeply processed. In this case, retrieval of any one trace is as efficient as under the different-question condition (see Rowe, 1973a, b) and retrieval of all traces is more efficient.

In summary, Experiment 4 found no support for the encoding variability hypothesis but endorsed the levels of processing hypothesis as an explanation of the spacing effect. It also supported the results of Rowe (1973a, b) and militated against the argument of Underwood (1971) that recognition and judgments of frequency involve basically the same process. The recognition results of Experiment 4 can be interpreted entirely within the levels of processing framework, i.e., greater depth of processing accounts for the advantage of both long-lag repetitions over short-lag repetitions and of different encoding contexts over similar encoding contexts. The judgments of frequency require explanation in terms of both levels of processing and variable encoding contexts. The levels of processing hypothesis accounts for the effect of spacing while the differences between the two question conditions reflect additionally the difficulty of retrieving memory traces of variably encoded stimuli.

Experiment 5

The levels of processing hypothesis states that there are two basic memory retrieval processes---a reconstructive process which is used following longer retention intervals and a rapid scanning which is efficient over short retention intervals. An item presented two or more times with long spacings is assumed to involve at each repetition a deep, reconstructive process which enhances retention. On the other hand, the occurrence of a repetition after a short spacing interval is assumed to initiate the scanning process as a subject attempts to retrieve the trace of the initial presentation. This scanning process is further assumed to be rapid and shallow and therefore the occurrence of a repetition after a short lag will not add much to the "strength" of the memory trace of the repeated item. Thus items repeated with short spacings are poorly retained and so the typical spacing effect is produced.

The purpose of experiment 5 is to obtain data relevant to the sense in which the scanning process used with short-lag repetitions is "shallow". For instance, Lockhart (Note 1) suggests that "superficial, non-semantic cues may suffice" to retrieve the trace of the initial presentation when a repetition is presented after a short spacing. If this suggestion is correct, the scanning process would be closely related to the processing required when subjects are engaged in a graphemic/orthographic incidental learning task. The question asked in this experiment then is whether the term "shallow" as applied to the scanning process is consonant with the term "shallow" as applied to the processing of physical information.

In order to answer this question, Experiment 4 was repeated with a new sample of subjects who were asked to answer questions concerning the physical properties of experimental words. Following the view of Lockhart et al, this task should induce shallow encoding processes which would lead to poor retention relative to the results from Experiment 4. Furthermore, if the scanning of recent memory and the processing of graphemic/orthographic features occur at the same shallow level, then the two question conditions will produce very similar results and neither will show any effect of spacing. The disappearance of a spacing effect under these conditions should arise from two causes. First, recognition of repetitions _as_ repetitions should be relatively poor at longer lags with a shallow task and this should attenuate the effect of spacing. Secondly, the reconstructive process will be producing only shallow re-encodings when it is successfully applied to repetitions. Thus, both massed and spaced repetitions, whether with the same or different questions, will receive shallow encodings. Rose and Rowe (1976, Experiment 2) found no effect of lag with their coding group who carried out a task demanding physical analyses of words. However, they did not manipulate the context of the words as was done here.

Method

Materials. The materials used here were exactly the same words distributed in exactly the same manner in six packs of cards as in Experiment 4. The difference in the two experiments lay only in the types of questions, which in this experiment concerned physical attributes of the words such as type of print, color of ink, numbers of syllables, letters, consonants, vowels, etc. (See Table 8, page 81, for examples of the types of questions asked here.) Unfortunately, the number of different questions which can be asked about the physical properties of words is much more restricted than is the case with semantic properties. As a result, each word could not have its own set of unique questions. Slight variations of 16 different questions were used and allotted quasi-randomly with the proviso that words occupying neighboring ordinal positions were not given the same question.

In order to make the incidental task meaningful, half of the words were typewritten and half were handwritten. Orthogonal to these factors, half of the words were in capital letters (or capitalized) and half were in small letters. In addition, the typewritten words could be either italic or "ordinary" type while the handwritten words could be either written or printed and in any one of red, blue, or black inks.

Subjects. The subjects were 49 undergraduates of Memorial University, of whom the results of one were discarded for failure to obey instructions. The subjects were paid $3.00 each and were tested in the same fashion as in Experiment 4.

Procedure. The experiment was carried out in exactly the same manner as Experiment 4. In order to lend credibility to the incidental task, which might appear to be pointless due to the "obvious" answers of most questions, the subjects were told that their behavior on the task would be compared to that of a group who answered semantic questions concerning the words in the packs.

Results

An examination of the responses on the incidental task revealed that one subject left several blanks on his response sheet, so his responses on the final judgment test were not counted. The results for Experiment 5 were scored and analyzed like those for Experiment 4. The results of the secondary analyses will be mentioned now and not discussed further.

With regard to the positive versus negative type of question under the different-question condition, it was found that those words asked an "extra" positive question (at frequencies of 3 and 5) were given slightly higher judgments than those words asked an "extra" negative question. The means were 2.56 and 2.49 respectively. Spacing did not appear to have any differential effect. Turning to the same-question condition, one can see the spacing functions for positive and negative questions collapsed across frequency in Figure 13. Unlike Experiment 4, the mean judgments of frequency for these two sub-conditions were considerably different at 2.95 and 2.38 for positive and negative questions respectively, $F(1,47) = 21.54$. However, the general consequence of spacing was the same in each case as indicated by the non-significant effects of any interaction involving the factor of question type.

Judgments of Frequency. The mean judgment for each frequency x spacing cell within each condition is shown in

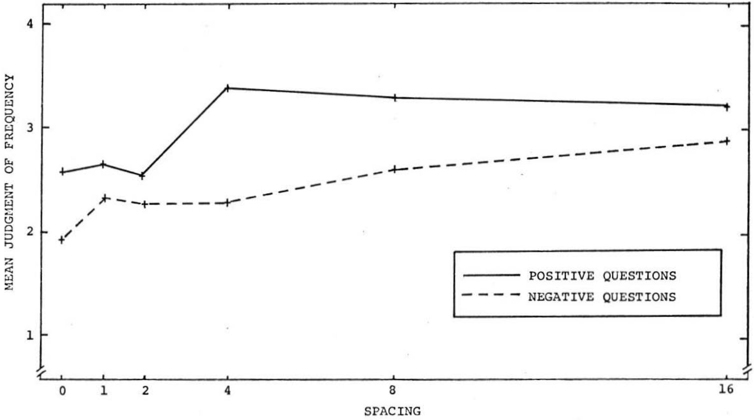

Figure 13. Mean judgments of frequency for condition S by type of question and spacing, collapsed across presentation frequency (Experiment 5).

Table 10 and depicted in Figures 14 and 15. The figures indicate the effect of spacing is negligible with different questions but present in the same-question condition, although not as prominent as in the same type of condition in Experiment 4. An analysis of variance with frequency, lag, and question condition as factors was performed and summarized in Table I. The analysis confirmed that the mean judgment for the same-question condition was significantly greater than the mean judgment for the different-question condition, $\underline{F}(1,47) = 19.16$. In addition, there was a significant effect of frequency, $\underline{F}(2,94) = 82.63$, a significant effect of spacing, $\underline{F}(5,235) = 2.64$, $\underline{p} < .025$, a significant question condition x spacing interaction, $\underline{F}(5,235) = 4.06$, $\underline{p} < .01$, which reflects the differential effect of spacing upon the two question conditions as described above, and a significant frequency x question condition interaction, $\underline{F}(2,94) = 3.22$, $\underline{p} < .05$, which indicates that frequency had a greater effect with same question than with different questions.

Because the triple interaction was also significant, $\underline{F}(10,470) = 2.81$, $\underline{p} < .01$, separate two-way analyses for each level of presentation frequency were carried out. At frequencies of 2 and 3, only question type showed a significant effect, $\underline{F}$'s$(1,94) = 4.27$ and 6.75 respectively, $\underline{p}$'s $< .05$. At frequency of 5 there was a significant effect of question type, $\underline{F}(1,94) = 32.95$, and a significant question condition x lag interaction, $\underline{F}(5,470) = 8.47$. Further

TABLE 10

Mean Judgments of Frequency by Question Condition
(Experiment 5)

	Spacing						
Frequency	0	1	2	4	8	16	$\bar{X}$
(a) Different Questions							
2	1.51	1.94	1.61	1.65	1.86	1.89	1.74
3	2.10	2.03	2.39	2.45	2.17	2.06	2.20
5	3.27	2.73	3.05	2.55	3.01	2.49	2.85
$\bar{X}$	2.29	2.23	2.35	2.22	2.35	2.15	2.26
(b) Same Questions							
2	2.07	1.99	1.80	1.83	2.21	2.09	2.00
3	2.02	2.23	2.60	2.65	2.77	2.86	2.52
5	2.70	3.21	3.01	4.11	4.04	4.31	3.56
$\bar{X}$	2.26	2.48	2.47	2.86	3.01	3.09	2.70

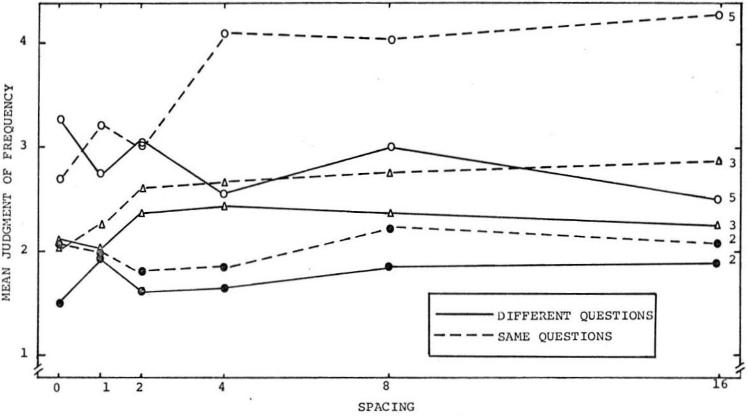

Figure 14. Mean judgment of frequency as a function of question condition, presentation frequency (2, 3, 5) and spacing (Experiment 5).

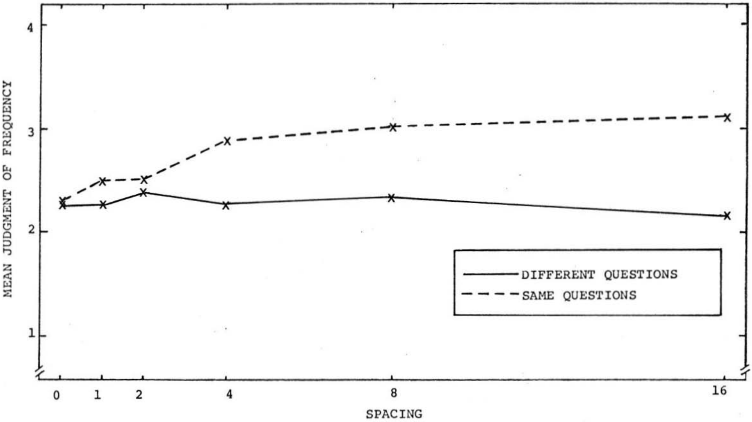

Figure 15. Mean judgment of frequency by question condition and spacing collapsed across presentation frequency (Experiment 5).

analyses showed for thrice-presented words under the same-question condition an effect of spacing of borderline significance, $\underline{F}(5,470) = 2.34$, $\underline{p} = .05$. Spacing also had a significant effect for words presented 5 times with the same questions, $\underline{F}(5,470) = 8.89$ where mean judgments at lags of 4, 8, and 16 exceeded mean judgments at lags 0, 1, and 2. This results can be seen quite clearly in Figure 14 which shows the mean judgments at each level of frequency by each question condition x frequency condition. None of these was significant except for the same-question condition, which had a significant linear component, $\underline{F}(1,470) = 6.61$, $\underline{p} < .025$, at frequency of 3 and significant linear and quadratic components, $\underline{F}(1,470) > 8.51$, $\underline{p}$'s $< .01$, at frequency of 5.

Recognition Scores. As in Experiment 4, derived recognition scores were calculated for each question condition and analyzed. The probabilities of correct recognition collapsed across frequency are shown in Figure 16. The results here are similar in form to the recognition scores in Experiment 4 with different questions producing better recognition than same questions while the latter shows more effect of spacing. The analysis of variance confirmed this similarity with frequency and question condition once more producing the only significant effects. For question condition, $\underline{F}(1,47) = 5.83$, $\underline{p} < .025$. The effect of spacing was once again only suggestive in the same-question condition. However, too much importance should not be attached to this null effect. Derived recognition scores are obviously less sensitive than judgments

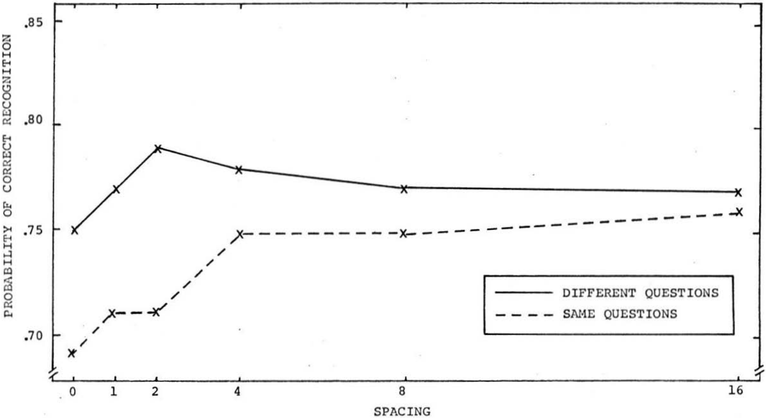

Figure 16. Mean probability of correct recognition as a function of question condition and spacing (Experiment 5).

of frequency. If there were more observations per frequency x spacing cell than the 96 used here, the effect of spacing under the same question condition might very have been significant.

The recognition scores, like the judgments of frequency, were considerably lower in Experiment 5 than in Experiment 4. This can be seen, not only in Figures 11 and 16, but also in the recognition error rates. For Experiment 4, these were 0.20 for false alarms, 0.04 for misses, and 0.06 for both types of error combined. The corresponding rates for Experiment 5 were 0.26, 0.29, and 0.28. Obviously the incidental task used in Experiment 5 led to a much higher rate of failure to recognize the experimental words on the final test.

The recognition scores for the same-question condition were also examined for the effects of positive questions versus negative questions. The only significant factors in the analysis of variance were presentation frequency, $F(2,94) = 7.48$, and type of question, $F(1,47) = 11.48$, both p's $< .01$. Once again, recognition of words presented with negative questions was worse at a hit rate of 0.69 than the recognition of words presented with positive questions, whose hit rate was 0.76. The effect of spacing, such as it was, was similar for positive and negative questions, insofar as recognition performance at higher levels of lag tended to exceed that at lower levels.

Conditional Judgments of Frequency. Judgments of frequency conditional upon correct recognition were again calculated for each question condition x frequency x spacing cell. These data were considered to be especially pertinent in Experiment 5 where the recognition performance was considerably poorer than in Experiment 4. However, as can be seen from Figure 17, the pattern of the condition judgments was similar to that for the unconditional judgments. As in Experiment 4, the judgments with repeated questions increased relative to the different-question condition when judgments were conditional upon correct recognition. This tendency can be seen clearly by comparing Figures 15 and 18.

Results at Frequency of 1 versus Results at Lag of 0. As in Experiment 4, the judgments of frequency and recognition hits for once-presented items were compared against the massed repetitions of items under the same-question condition. The mean judgments for the massed repetitions at each level of frequency can be seen in Table 10. The mean judged frequency of the once-presented (and tested) words was 1.18. An analysis of variance showed a significant difference among the means, $\underline{F}(3,141) = 10.20$, which was due to the mean of once-presented words being significantly less than the other three and to the mean at frequency of 5 being significantly larger than the others.

Turning to the recognition scores, the means were again significantly different, $\underline{F}(3,141) = 5.03$, $\underline{p} < .01$.

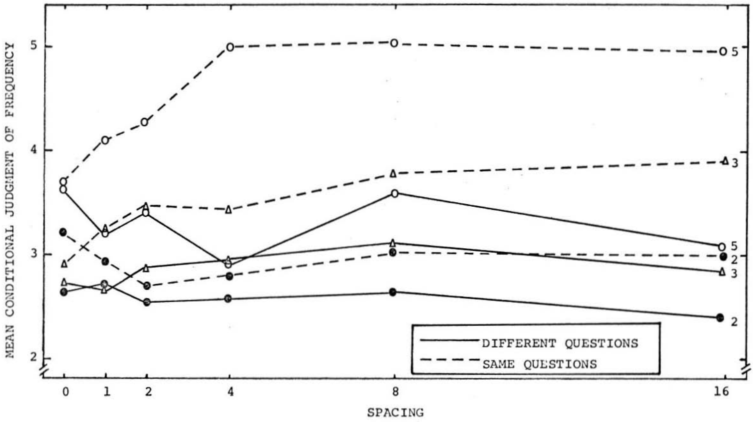

Figure 17. Mean judgment of frequency as a function of question condition, presentation frequency (2, 3, 5), and spacing, conditional upon correct recognition (Experiment 5).

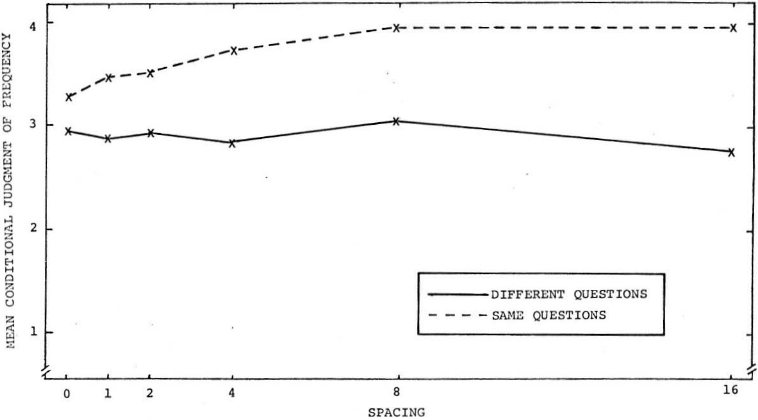

Figure 18. Mean judgment of frequency by question condition and spacing, conditional upon correct recognition (Experiment 5).

This difference was due to the probability of correct recognition of once-presented words, which was 0.51, being significantly lower than the other three probabilities. The results here then are consonant with those of Experiment 4, insofar as they show that repetitions at the same (or even lower) level of encoding as the first presentation do enhance the long-term memory of the repeated item. They also replicate the findings of Nelson (1977, Experiment 2) that this enhancement can occur even when a "shallow" incidental task is used on the study trial.

Discussion

The results of Experiment 5 are similar in form, although not in magnitude, to the results of Experiment 4. In both experiments the different-question condition exceeded the same-question condition on measures of recognition, especially at small values of lag. The reverse was generally true for judgments of frequency, especially at higher values of lag. When judgments of frequency were measured, there was an attenuated effect of spacing with different question in both experiments. There was also a significant effect of spacing with same questions, although in Experiment 5 the effect was limited to items occuring three and five times. For measures of recognition, both experiments showed a tendency for performance to increase over spacing under the same-question condition, but in neither case was this tendency significant.

The major differences between Experiments 4 and 5 lie in the overall level of performance and the question condition x spacing interactions found with the judgments of frequency. Both measures of retention are considerably lower in Experiment 5 than in Experiment 4. This result replicates the findings of many other experiments (e.g., Craik & Tulving, 1975; Lockhart et al., 1976; Nelson, 1977) which show that long-term memory following a semantic incidental task exceeds long-term memory following a graphemic/orthographic task. Regarding the interactions, two points may be made. First the effect of spacing was smaller in Experiment 5 than in Experiment 4. Each level of frequency in Experiment 4 showed a significant question condition x spacing interaction. In Experiment 5, there was a significant interaction only at frequency of 5, although judgments of thrice-presented words with repeated questions showed a significant linear trend to increase over spacing. This diminished effect of spacing is no doubt a concomitant of the overall level of performance in the fifth experiment, i.e., each repetition at a long lag under same questions adds less to the "strength" of the memory trace in Experiment 5 than it does in Experiment 4. Secondly, the same question condition exceeds the different-question condition at all levels of spacing in Experiment 5. This finding indicates that repetition of a word accompanied by different graphemic/orthographic questions does not induce any greater depth of processing than does repetition with the same question at small values of lag. This indication is in contrast to Experiment 4 where different semantic questions did apparently increase the relative depth of processing of massed repetitions such that the difficulty of retrieving traces under the different-question condition was offset.

The main concern of Experiment 5 was the type of "shallow" encoding which the scanning process of memory retrieval was assumed to produce. The predictions of no differences in mean judgments of frequency for the two question conditions and of no effect of spacing were clearly not supported. The same-question condition exceeded the different-question condition and also showed significant trends to increase over spacing. This spacing effect indicates that the reconstructive process can be efficiently applied at longer lags even when the reconstructed event concerns the physical properties and not the meaning of the word. This view in turn is consonant with the suggestion by Lockhart et al. (1976, p. 78) that the concept "depth of processing" may be used in two different senses, the sense of "domain" (see Sutherland, 1968; 1972) and the sense of elaboration of the processing carried out within a domain. Following this suggestion, the argument develops that the scanning and reconstructive retrieval processes occur within a particular domain, with reconstruction being more elaborate than scanning. On the other hand, the different tasks used in Experiments 4 and 5, semantic as opposed to physical, would involve separate domains. Thus the scanning-reconstruction dimension is considered to be independent of the physical-semantic dimension. This view accounts for the similarity in the pattern of the effects of spacing in Experiments 4 and 5 and the dissimilarity in the levels of memory performance.

The arguments of the preceding paragraph account for the results of the last two experiments where the effects of the type of task and spacing are concerned. Turning to the effect of the encoding context, it was argued preceding Experiment 4 that the different question condition would force subjects to process repetitions to a fairly deep level, since mere scanning would be precluded. Because this argument was supported by the recognition results but not by the judgments of frequency, it was amended to include the factor of retrievability. Specifically, the variable encoding contexts of repetitions not only induced deep processing via elaboration of the encoding processes within a domain but they also provided variable retrieval cues. These variable cues are no hindrance (and possibly an advantage) when only one memory trace is required as in recognition but they inhibit the retrieval of all memory traces of a repeated item which is required in making judgments of frequency.

In summary, the results of Experiment 5 supported and extended Experiment 4 in arguing for the levels of processing hypothesis as an explanation of the spacing effect and the role of encoding context in forming judgments of frequency. In addition the data from Experiment 5 indicated that the term "level of processing" as applied to the spacing effect differs in sense from the "level of processing" defined by a task.

GENERAL DISCUSSION

The experiments reported here were directed towards finding an explanation for the effect on memory of the spacing of repetitions. The research was directed particularly toward an explanation which would be consistent with data from studies of judgments of frequency. However, since the spacing function takes a similar form across a range of measures of retention, any hypothesis relevant to the effect of spacing upon judgments of frequency should be applicable to other tasks as well.

The experiments investigated the spacing function up to the point of asymptote or beginning of decline in retention at high values of lag. They did not concern the overlap of encoding and test contexts nor the effects of various test retention intervals. As such then, the data and the levels of processing hypothesis may be viewed as complementary to Glenberg's (1976) views concerning the occurrence of an asymptote. However, insofar as Glenberg invokes variable encoding contexts as an explanation of spacing <u>up to</u> the asymptote, the empirical data here do not support his explanation.

The results of Experiment 1 suggested that the spacing effect arose from some difference between continuous and terminal judgments of frequency. This difference could have arisen from the use of different strategies for the two types of judgment or from some differential loss of information over spacing during the retention interval preceding the terminal test. The effect of spacing upon the bias-free discrimination coefficients in the first experiment contradicted some results of Hintzman (1969). He found only a massed presentations effect when another bias-free measure, a forced-choice test, was used. Experiment 2 failed to replicate Hintzman's results when a more sensitive test was used and hence found further evidence for the dismissal of bias factors. Experiment 3 found no support for the view that test strategy was an explanatory factor. Taken together, the results of the first three experiments support the conclusion that the spacing effect is due to differential forgetting over lag during a relatively long retention interval.

Experiments 4 and 5 extended this view by contrasting the encoding variability and levels of processing hypotheses as explanations of the differential forgetting. The encoding variability hypothesis ascribes the spacing effect to the increasing variability in the encoding contexts of repeated items as the spacing between the repetitions increases. The levels of processing hypothesis attributes the effect to the progressively deeper processing over lag of repeated items. The results provided no support for the encoding variability hypothesis but were consistent with the levels of processing view of the spacing effect.

Taken together, Experiments 4 and 5 showed the effects of three critical factors. First, the type of task (semantic versus graphemic/orthographic) determined the overall level of performance within each experiment. This factor established the level of processing in the sense of the domain of the processing. The second factor was the spacing between repetitions which, for the same-question condition in each experiment, was directly related to performance. This result was considered to reflect the level of processing in the sense of degree of elaboration of processing within a domain. Thirdly, the encoding context produced different patterns of results with recognition hits and judgments of frequency. The effect of this factor was explained in terms of an interaction between the demands of the particular memory test and the retrieval cues provided by the encoding context. The introduction of the effects of all three factors into an explanation of the results produces unexpected complexity but is nonetheless demanded by the data. From a positive point of view, the interaction between levels of processing and contextual cues may be considered as a step towards bridging encoding and retrieval factors. However, it should be noted that the arguments derived from the last two experiments apply only when encoding contexts are manipulated and test contexts are not. Where both encoding and test contexts are manipulated, different arguments may prevail (see Thomson & Tulving, 1970; Tulving & Thomson, 1973).

The arguments put forward to explain the results of Experiments 4 and 5 are based upon certain assumptions concerning the form of memory and have wider implications for research. One major assumption is that subjects who meet a repeated stimulus in a sequence attempt to re-instate the original encoding of that stimulus. The main problem with this assumption is that there has been little direct study of the "looking-back" process, although the experiments by Jacoby (1974) are a notable exception. Other exceptions which are particularly appropriate to the spacing effect are the studies of Hintzman and Block (1973) and Hintzman, Summers, and Block (1975a). They found that subjects on a terminal test could judge quite accurately the spacing between two presentations of a word in a long sequence or between a word and a common associate of that word. However, subjects showed no ability to judge the spacing between two unrelated words. Hintzman and his colleagues interpret these results as suggesting that when a repetition (or a strong associate) of a word occurs, the trace of the initial presentation is implicitly retrieved and its recency encoded. Given that the recency of the previous presentation is encoded with the current presentation, then it is reasonable to assume that the total frequency of occurrence of the word is encoded as well. This encoding of frequency information has already been suggested by Begg and Rowe (1972) and Hasher and Chromiak (1977).

If information concerning frequency of occurrence and spacings is encoded at the last presentation of an item, one may then ask why spacing affects retention on a later test? All one would have to do on the test is retrieve the relevant information from the last presentation, a process which would be a function of retention interval but not of spacing. Experiments 1 and 3 show that subjects do not do this even when the continuous task explicitly requires them to form judgments of frequency. The length of the retention interval is probably the key to this dilemma. When the retention interval is relatively long, the frequency information at the last presentation may be in a degraded form. Alternatively, as Glenberg (1976) suggests, the test item after a long retention interval may tend to retrieve the traces of any (or all) occurrences of the item in the sequence. Since one cannot rely upon retrieving the trace of the last presentation after a long retention interval, one cannot have confidence in the encoded frequency information which is retrieved at the time of testing. In either event, a check of the traces of previous occurrences would have to be made and spaced repetitions would have traces which are more likely to be retrieved due to deeper processing. The form of memory which is pre-supposed then by these arguments is an associative one, in which the attributes of previous presentations and associates of a word are automatically retrieved when the word is presented but whose separate traces are nonetheless retained as a form of redundancy to be used after long retention intervals.

The levels of processing hypothesis emerges from this set of experiments as the best current framework for accomodating what is known of the spacing effect (up to the asymptote). It suffers however, from a major weakness in that there is no direct evidence for the existence of the scanning and reconstructive processes. For instance, the scanning versus reconstructive dichotomy is not reflected in the spacing function, i.e., there appears no point where one process seems to replace the other. If it exists, then one should be able to force subjects to use one or the other process throughout the sequence and eliminate the effect of spacing. However, attempts to manipulate these processes would appear to be fruitless until more is known about their nature.

Alternatively, it may be beneficial to consider the scanning and reconstructive processes as vague areas at either end of an continuum. As the matter now stands, one would have to postulate that, as spacing increases, subjects progressively use the reconstructive process in preference to the scanning of recent memory. A more fruitful approach may be to replace the dual process view with a concept like effort which is postulated to increase gradually (within a domain) for the processing of a repetition as the "strength" of the previous presentation fades. Kahneman (1973) suggests that the notion of attention be conceived in just such terms. The previously mentioned study by Johnston and

Uhl (1976), who used a subsidiary reaction time task with massed and distributed repetitions, is a step in this direction. A useful project for further research would be a replication of this study with several levels of spacing instead of merely massed repetitions versus distributed repetitions. Such a study should determine whether the processing of a repetition requires more effort as spacing increases. If this outcome were found, the levels of processing explanation of the spacing effect could then be considered as a form of involuntary attention hypothesis.

In summary, the findings of this thesis support the view that the effect of spacing upon the retention of a repeated item arises from the increasing depth of processing of a repetition over spacing. Furthermore, this depth of processing is to be considered in the sense of elaboration of processing within a domain, as shown by the qualitative similarity of the spacing effects with semantic and non-semantic tasks in conjunction with the quantitative differences in levels of retention usually found with such tasks. Finally, evidence was found that the level of processing interacts with the effect of context on the retrievability of traces and that a judgment of frequency task can be differentiated from a recognition task on the basis of this effect.

REFERENCES

Anderson, J.R., & Bower, G.H. Recognition and retrieval processes in free recall. Psychological Review, 1972, 2, 97-123.

Begg, I. Estimation of word frequency in continuous and discrete tasks. Journal of Experimental Psychology, 1974, 102, 1046-1052.

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

Bellezza, F.S., Winkler, H.B., & Andrasik, F. Jr. Encoding processes and the spacing effect. Memory and Cognition, 1975, 3, 451-457.

Bevan, W., Dukes, W.F., & Avant, L.L. The effect of variation in specific stimuli on memory for their superordinates. American Journal of Psychology, 1966, 79, 250-257.

Bjork, R.A., & Allen, T.W. The spacing effect: Consolidation or differential encoding? Journal of Verbal Learning and Verbal Behavior, 1970, 9, 567-572.

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

Craik, F.I.M., & Tulving, E. Depth of processing and the retention of words in episodic memory. Journal of Experimental Psychology: General, 1975, 104, 268-294.

D'Agostino, P.R., & De Remer, P. Item repetition in free and cued recall. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 54-58.

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

Elmes, D.G., Greener, W.I., & Wilkinson, W.C. Free recall of items presented after massed- and distributed-practice items. American Journal of Psychology, 1972, 85, 237-240.

Flexser, A.J., & Bower G.H. Further evidence regarding instructional effects on frequency judgments. Bulletin of the Psychonomic Society, 1975, 6, 321-324.

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 Behavior, 1972, 11, 801-808.

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

Glenberg, A. Monotonic and nonmonotonic lag effects in paired-associate and recognition memory paradigms. *Journal of Verbal Learning and Verbal Behavior*, 1976, *15*, 1-16.

Hasher, L., & Chromiak, W. The processing of frequency information: An automatic mechanism? *Journal of Verbal Learning and Verbal Behavior*, 1977, *16*, 173-184.

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 in Cognitive Psychology: The Loyola Symposium*. Potomac, Md.: Lawrence Erlbaum Associates, 1974.

Hintzman, D.L. Repetition and memory. In G.H. Bower (Ed.), *The Psychology of Learning and Motivation*, Vol. 10. New York: Academic Press, 1976.

Hintzman, D.L., & Block, R.A. Memory for the spacing of repetitions. *Journal of Experimental Psychology*, 1973, *99*, 70-74.

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 Behavior*, 1973, *12*, 229-239.

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

Hintzman, D.L., Summers, J.J., & Block, R.A. Spacing judgments as an index of study-phase retrieval. Journal of Experimental Psychology: Human Learning and Memory, 1975, 1, 21-40. (a)

Hintzman, D.L., Summers, J.J., & Block, R.A. What causes the spacing effect? Some effects of repetition, duration and spacing on memory for pictures. Memory and Cognition, 1975, 3, 287-294. (b)

Hintzman, D.L., Summers, J.J., Eki, N.T., & Moore, M.D. Voluntary attention and the spacing effect. Memory and Cognition, 1975, 3, 576-580.

Jacoby, L.L. The role of mental contingthe in memory: Registration and retrieval effects. Journal of Verbal Learning and Verbal Behavior, 1974, 13, 483-496.

Johnston, W.A., & Uhl, C.N. The contributions of encoding effort and variability to the spacing effect on free recall. Journal of Experimental Psychology: Human Learning and Memory, 1976, 2, 153-160.

Kahneman, D. Attention and Effort. Englewood Cliffs, New Jersy: Prentice-Hall, Inc., 1973.

Landauer, T.K. Reinforcement as consolidation. Psychological Review, 1969, 76, 82-96.

Lockhart, R.S., Craik, F.I.M., & Jacoby, L.L. Depth of processing in recognition and recall: Some aspects of a general memory system. In J. Brown (Ed.), Recognition and Recall. London: Wiley, 1976.

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

Martin, E. Stimulus meaningfulness and paired-associate transfer: An encoding-variability hypothesis. Psychological Review, 1968, 75, 421-441.

Martin, E. Stimulus encoding in learning and transfer. In A.W. Melton & E. Martin (Eds.), Coding Processes in Human Memory. Washington: Winston, 1972.

Melton, A.W. Repetition and retrieval from memory. Science, 1967, 158, 532.

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

Nelson, T.O. Repetition and depth of processing. Journal of Verbal Learning and Verbal Behavior, 1977, 16, 151-171.

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

Paivio, A.U. Spacing of repetitions in the incidental and intentional free recall of pictures and words. Journal of Verbal Learning and Verbal Behavior, 1974, 13, 497-511.

Paivio, A.U. Coding distinctions and repetition effects in memory. In G.H. Bower (Ed.), The Psychology of Learning and Motivation, Vol. 9. New York: Academic Press, 1975.

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.

Robbins, D., & Wise, P.S. Encoding variability and imagery: Evidence a spacing-type effect without spacing. Journal of Experimental Psychology, 1972, 95, 229-230.

Rose, R.J., & Rowe, E.J. Effects of orienting task and spacing of repetitions on frequency judgments. Journal of Experimental Psychology: Human Learning and Memory, 1976, 2, 142-152.

Rowe, E.J. Context effects in judgment of frequency. Bulletin of the Psychonomic Society, 1973, 2, 231-232. (a)

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

Rowe, E.J., & Rose, R.J. Effects of orienting task, spacing of repetitions, and list context on judgments of frequency. Memory and Cognition, 1977, 5, 505-512.

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

Schwartz, M. The effect of constant vs. varied encoding and massed vs. distributed presentations on recall of paired associates. Memory and Cognition, 1975, 3, 390-394.

Shaughnessy, J.J. Persistence of the spacing effect in free recall under varying incidental learning conditions. Memory and Cognition, 1976, 4, 396-377.

Shaughnessy, J.J., Zimmerman, J., & Underwood, B.J. Further evidence on the MP-DP effect in free recall learning. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 1-12.

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

Sutherland, N.S. Outlines of a theory of pattern recognition in animals and man. Proceedings of the Royal Society, Series B, 1968, 171, 297-317.

Sutherland, N.S. Object recognition. In E.C. Carterette & M.P. Friedman (Eds.), Handbook of Perception, Vol. 3. New York: Academic Press, 1973.

Thomson, D.M. & Tulving, E. Associative encoding and retrieval: Weak and strong cues. Journal of Experimental Psychology, 1970, 86, 255-262.

Thorndike, E.L., & Lorge, I. The Teacher's Word Book of 30,000 Words. New York: Columbia University, Teacher's College, Bureau of Publications, 1944.

Tulving, E., & Thomson, D.M. Encoding specificity and retrieval processes in episodic memory. Psychological Review, 1973, 80, 352-373.

Tzeng, O.J.L. Stimulus meaningfulness, encoding variability, and the spacing effect. Journal of Experimental Psychology, 1973, 99, 162-166.

Underwood, E.J. Some correlates of item repetitions in free-recall learning. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 83-94.

Underwood, B.J. A breakdown of the total-time law in free-recall learning. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 573-580.

Underwood, B.J. Recognition memory. In J.T. Spence and H.H. Kendler (Eds.), Essays in Neobehaviorism: A Memorial Volume to Kenneth W. Spence. New York: Appleton-Century-Crofts, 1971.

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.

Wells, E., & Kirsner, K. Repetition between and within modalities in free recall. Bulletin of the Psychonomic Society, 1974, 2, 395-397.

Wickelgren, W.A., & Berian, K.M. Dual trace theory and consolidation of long-term memory. Journal of Mathematical Psychology, 1971, 8, 404-417.

Zimmerman, J. Free recall after self-paced study: A test of the attention explanation of the spacing effect. American Journal of Psychology, 1975, 88, 277-291.

REFERENCE NOTE

1. Lockhart, R.S. The spacing effect in free recall. Paper presented at the fourteenth annual meeting of the Psychonomic Society, St. Louis, November, 1973.

APPENDIX A

Summaries of Analyses of Variance

TABLE A

Experiment 1: Summary of Analysis of Variance of Terminal JoF Results

Source	df	MS	F
(a) Two-way analysis			
Subjects (S)	23	14.8012	
Presentation frequency (F)	2	66.1744	75.21**
F x S	46	.8799	
Spacing (L)	5	8.2777	15.23**
L x S	115	.5434	
F x L	10	2.3564	7.70**
F x L x S	230	.5009	

**$p < .001$

TABLE B

Experiment 1: Summary of Analysis of Variance of Final Continuous JoF Results

Source	df	MS	F
(a) Two-way analysis			
Subjects (S)	23	2.0436	
Presentation frequency (F)	2	289.1572	1072.72**
F x S	46	.2696	
Spacing (L)	5	1.4224	4.55**
L x S	115	.3128	
F x L	10	.6430	3.42**
F x L x S	230	.1878	

**$p < .001$

TABLE C

Experiment 1: Summary of Analysis of Variance of Final Continuous JoF's versus Terminal JoF's

Source	df	MS	F
Subjects (S)	23	12.1916	
Presentation Frequency (F)	2	314.4444	394.31**
F x S	46	.7975	
Spacing (L)	5	7.2632	14.72**
L x S	115	.4935	
Type of Test (T)	1	57.6580	12.42*
T x S	23	4.6428	
F x L	10	2.0055	5.02**
F x L x S	230	.3997	
F x T	2	39.3371	113.35**
F x T x S	46	.3470	
L x T	5	2.4766	6.77**
L x T x S	115	.3656	
F x L x T	10	1.0039	3.44**
F x L x T x S	230	.2922	

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

TABLE D

Experiment 1: Summary of Analysis of Variance of the Transformed Discrimination Coefficients

Source	df	MS	F
Subjects (S)	23	.1918	
Spacing (L)	5	.6844	8.04**
L x S	115	.0852	

**$p < .001$

TABLE E

(a) Experiment 2: Summary of Analysis of Variance of the Choices Made Within Pairs by Spacing

Source	df	MS	F
Subjects (S)	23	0.0000	
Spacing (L)	5	0.6330	14.21**
L x S	115	9.1899	

(b) Experiment 2: Summary of Analysis of Variance of the Choices Made Within Pairs by Spacing (Excluding L = 0)

Source	df	MS	F
S	23	0.0000	
L	4	39.2708	5.40**
L x S	92	7.2708	

**$p < .001$

TABLE F

Experiment 3: Summary of Analysis of Variance of the Final Continuous JoF's by Groups

Source	df	MS	F
Between Subjects			
Instructional group (G)	1	.2462	<1
Subjects within groups (S)	62	3.0396	
Within Subjects			
Presentation frequency (F)	2	425.6025	837.06**
G X F	2	.0707	<1
F x S	124	.5084	
Spacing (L)	3	.7045	1.94
G x L	3	.5433	1.49
L x S	186	.3639	
F x L	6	1.5997	8.83**
G x F x L	6	.1889	1.04
F x L x S	372	.1812	

**$p < .001$

TABLE G

Experiment 3: Summary of Analysis of Variance of the Terminal JoF's by Groups

Source	df	MS	F
Between Subjects			
Instructional group (G)	1	13.9483	2.15
Subjects within groups (S)	62	6.4855	
Within Subjects			
Presentation frequency (F)	2	159.5731	234.88**
G x F	2	.1809	<1
F x S	124	.6794	
Spacing (L)	3	30.9798	66.61**
G x L	3	1.0296	2.21
L x S	186	.4651	
F x L	6	4.4366	10.98**
G x F x L	6	.7535	1.87
F x L x S	372	.4040	

**$p < .001$

TABLE H

Experiment 4: Summary of Analysis of Variance of the Judgments of Frequency by Question Condition

Source	df	MS	F
Subjects (S)	47	29.3575	
Presentation frequency (F)	2	645.5625	173.29***
F x S	94	3.7253	
Question condition (Q)	1	3.6117	<1
Q x S	47	4.5945	
Spacing (L)	5	48.3733	21.27***
L x S	235	2.2740	
F x Q	2	11.0327	5.13**
F x Q x S	94	2.1489	
F x L	10	6.2759	4.21***
F x L x S	470	1.4895	
Q x L	5	18.0848	8.42***
Q x L x S	235	2.1472	
F x Q x L	10	3.2735	2.15*
F x Q x L x S	470	1.5250	

*$p < .025$
**$p < .01$
***$p < .001$

TABLE I

Experiment 5: Summary of Analysis of Variance of the Judgments of Frequency by Question Condition

Source	df	MS	F
Subjects (S)	47	24.2359	
Presentation frequency (F)	2	263.0850	82.63****
F x S	94	3.1840	
Question condition (Q)	1	80.2885	19.16****
Q x S	47	4.1915	
Spacing (L)	5	7.0398	2.64**
L x S	235	2.6696	
F x Q	2	8.7851	3.22*
F x Q x S	94	2.2251	
F x L	10	2.3369	<1
F x L x S	470	2.4376	
Q x L	5	10.2023	4.06***
Q x L x S	235	2.5110	
F x Q x L	10	6.9046	2.81***
F x Q x L x S	470	2.4555	

*$p < .05$
**$p < .025$
***$p < .01$
****$p < .001$