A TASK ANALYSIS OF LEXICAL DECISIONS

by

© CHARLES ARLO GORDON HAYMAN, B.Sc.

A thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Doctor of Philosophy

McMaster University

January 1983

A TASK ANALYSIS OF LEXICAL DECISIONS

DOCTOR OF PHILOSOPHY (1983)
(PSYCHOLOGY)

McMaster Univerisity
Hamilton, Ontario

TITLE: A Task Analysis of Lexical Decisions

AUTHOR: Charles Arlo Gordon Hayman B. Sc. (Acadia University)

SUPERVISOR: Professor L. L. Jacoby

NUMBER OF PAGES: xi, 150

Abstract

A task analyis of the Lexical Decisön Task (LDT) was performed. Several alternative explanations are explored, and experiments designed to investigate these models are reported. These experiments provide strong evidence for the use of multiple decision strategies in the LDT. Subjects responded to both visual and semantic attributes of targets.

Lexical decisions to repeated nonwords resulted in both increased errors and decision latencies. Apparently, the familiarity of specific visual graphemes serves as one basis for performance in the LDT. A basis that is independent of an item's true lexical status. Also, lexical decisions were biased by typography. This was interpreted as confirming the visual basis of information in the LDT.

Meaning in lexical decisions was studied by varying both the type of word referent (concrete or abstract), and the availability of meaning from a prior presentation. While meaning contributed to lexical decisions, its use depended upon both a stimuli's visual familiarity and the nature of the task demands.

It was concluded that the LDT does not measure a single process or memory structure (i.e., Lexical Memory), rather it reflects knowledge about the visual familiarity and the semantic uses of graphemes.

An additional topic was the nature of the psychological mechanisms supporting recognition. Several general models of retrieval are contrasted. It is found that criterion bias models describe both the accuracy and latency of lexical decisions better than models which assume an ordered search of memory.

An extension of the Random Walk decision process is presented. Within this model biases in decision often reflect differences in what is actually known about stimuli relative to a single criterion, rather than multiple and seperate decision biases located at seperate decision loci. The benefits of this analysis in explaining the effects of word frequency, stimulus repetition and typographic information in lexical decisions are discussed.

The completion of this thesis would not have been possible without the support and encouragement of several people. Drs. Larry Jacoby and Lee Brooks provided questions and suggestions that were both stimulating and challenging. I thank them for their many comments, and Larry especially for his (endless?) support. Also, I extend my gratitude to Rosemary Young for many years of friendship, and for her careful reading and useful comments about the first draft of this thesis.

In addition, I would like to acknowledge the contributions of the many graduate students and faculty, with whom I shared many informal conversations, and who provided an enriching and enthusiastic academic and social environment. I am especially indebted to Don Shattuck and Bruce Whittlesea for their support, patience, and suggestions during the writing of this thesis. Last but not least, I thank Dr. Steve Link for the ideas and encouragement he has given me during my stay at McMaster.

TABLE OF CONTENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES

The Lexical Decision Task has become a standard test for investigating visual word recognition, and has been used to provide evidence about the structure and method of accessing information from a hypothesized lexical memory. The Lexical Decision Task has been used to examine the development of elementary reading skills in children, and the loss of reading skills in some patient populations. While widely used, it is not clear what abilities are actually measured in the Lexical Decision Task. In light of its frequent and varied use, it is necessary to gain a clearer understanding of what the Lexical Decision Task can tell us about the structure and the use of memory. The first section of the paper will provide an overview of the current models of lexical decision, followed by a new task-relative interpretation of the Lexical Decision Task. In the second section, five experiments designed to investigate the influence of task variables on lexical decision will be described, and the results of those experiments will be related to the prior discussion of theories.

In a typical Lexical Decision experiment, subjects are required to make two-alternative, forced-choice decisions concerning whether a string of letters represents an English word or is a nonsense string. Usually it is assumed that subjects perform this task by responding yes only when they can identify the test string as a word, and by responding no when they cannot succeed in identifying the test string as a word. It seems obvious that at some level the subject must be comparing the test string with a memory, or memories. This comparison, however, may not invove a lexical memory in which all words that the person knows are represented.

Consider a Lexical Decision Task where the nonsense distractor is a string of non-letter symbols, for example - <u>#$%*$</u>. In this task, the subject could respond <u>nonword</u> by rejecting target strings composed of non-letters, and <u>word</u> by accepting strings with letters as words. Thus, in some situations, subjects could perform a Lexical Decision Task by rejecting nonword distractors, and accepting words on the basis of constituent familiarity. In this example, words are only indirectly recognized and are not directly identified. Consider, next, a Lexical Decision Task where the nonsense distractors take the form of orthographically illegal strings of letters, such as <u>rbikti</u>. Once more it is possible that nonword distractors of this type could be rejected on the basis of constituent familiarity (i.e., letter bigrams and trigrams). Moreover, if words were assessed mainly in terms of the presence of familiar bigrams and familiar trigrams, then here to words would be only indirectly recognized and not directly identified.

In the typical lexical decision experiment, the nonword distractors are orthographically legal and pronounceable spellings, that is, pseudowords. Because of this intentional similarity in structure with words, it is usually tacitly assumed that the only useful strategy available to a subject is to respond on the basis of an explicit word identification, or on the failure to achieve such an identification. It is usually assumed that the subject possesses in memory a specialized set of items, a lexical memory, that can guarantee that a spelling pattern is a meaningful linguistic unit; if the target string matches one of these marked items then it must be a word.

Despite the use of pseudowords as distractors, a major aim of the present paper will be to demonstrate that subjects frequently respond on bases other than the sheer "wordness" of the stimuli. Responses to both words and nonwords in a Lexical Decision Task can reflect differences in the visual familiarity of the target graphemes. A pseudoword is not only semantically meaningless, it is also structurally unfamiliar. In addition, a subject could potentially make a lexical decision on the basis of phonemic information, i.e., "Does it have a familiar name?"; and/or on the basis of semantic information, i.e., "Does it refer to anything in specific?". In general, it can be assumed that the more familiar a string's name is, and the richer its semantic associations are, the more likely it is that that string is a true word. Phonemic and semantic information, though, demand some form of preliminary visual processing; and it may be that they constrain lexical decisions by verifying and supporting the preliminary visual infomation. Hence non-visual information may be most important when visual structural information is uncertain, or when a subject is striving for accuracy. In short, lexical decisions may be multidimensional judgments, drawing information from diverse memory sources, if and when it is needed. There is no a priori reason to believe that subjects rely on either a single process, or a single information source for making the decision. Rather, it seems more likely that they will use any resource that is necessary to get the job done, with a preference for the simplest and easiest method.

In summary, there are at least two ways of explaining nonword rejection, as well as word acceptance in the Lexical Decision Task. One method, characterized by the lexical store models, explains performance by assuming success or failure in achieving a word identification in some lexical memory. An alternative method, a relative familiarity hypothesis, explains performance by assuming task relative judgments of stimulus familiarity, where high familiarity corresponds to a word while low familiarity corresponds to a nonword. This relative familiarity hypothesis assumes that the Lexical Decision

Task does not measure a unique and unitary skill (i.e., word recognition or lexical recognition), rather it measures a subject's ability to perform a memory based discrimination, with words as targets and nonwords as distractors. Moreover, this memory based discrimination may use multiple distinctions such as visual, phonemic, and semantic knowledge. During the course of this paper I hope to demonstrate the benefits of viewing lexical decision as a relative familiarity judgment, rather than as involving a word (lexical) identification.

The relative familiarity hypothesis assumes that there is not a specific set of detectors in memory which can guarantee that a word is present. Rather, it assumes that there are memories about things we have experienced and call words, memories about what they look like, what they imply, and how they can be used. To the extent that a pseudoword evokes similar memories, it is likely to be confused with a word. The identification of words and rejection of psudowords can occur at one or several of these discriptive levels, and will depend upon the type of attributes pseudowords elicit and fail to elicit in memory.

Theories of Lexical Decision

The two most influential theories of the Lexical Decision Task, Morton's Logogen System and Becker's Verification Model, are lexical store models. Common to both of these theories is the assumption that there exists a "dictionary-like" lexicon that is separate from the rest of memory. Subjects are said to base their lexical decisions on the results of attempts to find a match between a target stimulus and members of this internal lexicon. The two models differ, however, in terms of their primary dependent variable. The Logogen System was designed to describe probability correct in the Word Detection Task and has been generalized to the Lexical Decision Task. The

Verification Model is specific to the Lexical Decision Task and is primarily concerned with decision latency. A third theory introduced here, Relative Familiarity Judgment, differs from the earlier models in that a separate lexicon is not postulated. Lexical decisions are seen as using memory as a whole, and being based on a variety of types of information. Furthermore, the theory incorporates both response accuracy and response latency, and specifies the relationship between these two measures. The following section describes the relevant assumptions for each of the three theories, and concludes by discussing the differences between them.

Morton's Logogen System

Morton (1969) postulated a set of decision units, logogens, which represented words. For each word there was a corresponding logogen, which served as the connection between sensory information and a verbal response. The logogens operated independently and in parallel to detect a probable match between the current sensory input and the featural information contained in the logogen.

The probabalistic nature of logogens reflected Morton's wish to explain the effect that variables such as word frequency, repetition and semantic context had on word detection in terms of a passive decision mechanism. This passive bias was seen by Morton and Broadbent (1967) as a preset difference in response thresholds (a Response Bias) and was contrasted with theories which proposed an attentional change in visual encoding and/or memory comparison (an Encoding Bias). As each logogen had a separate response threshold and operated in a manner similar to detectors in detection theory (Green & Swets,1966), the Logogen System provided a passive bias for detection.

Activation of a Logogen. In the absence of a stimulus, all logogens receive variable and random information, with a variance of one and a mean of zero. In the presence of a stimulus, logogens receive information in direct proportion to the similarity between the encoded stimulus and the features specified in the logogen. When the information accumulated by a logogen exceeds a threshold value, a logogen will signal the presence of its target word by firing.

The Word Frequency Effect and the effect of a repetition were attributed to differences in logogen thresholds. A logogen's threshold value decreased as the frequency of its word in the language increased, and as a function of repetition. Since logogens for repeated and frequently seen words have lower thresholds than logogens for infrequently seen words, then these low threshold logogens will typically require less stimulus information before firing.

There is a cost associated with the use of low threshold logogens, however, as low threshold logogens have an increased probability of being activated by a similar, but inappropriate stimulus. Consequently, the benefits found for frequent and repeated words are offset, in theory, by the cost of incorrectly producing these items as false alarms (Broadbent, 1971). While this cost/benefit relationship was an empirical fact Morton felt he had to explain (see Broadbent, 1967), it is an important and fundamental feature of the Logogen System. In the Logogen Systems there is no preferential uptake of information for frequent or repeated words; in the terms of Signal Detection theory, there is not an increase in <u>d'</u>, there is only an increase in <u>Beta</u>. This is a major difference between the Logogen System and the Relative Familiariy Judgment model, which predicts an increase in both terms, and will be discussed in greater detail in a later section.

A final feature of the Logogen System is that the quality of information available to a logogen was assumed to be the sole property of the stimulus presentation (duration, contrast, legibility, etc.), and was held to be independent of the differences in logogen threshold. Consequently, variations in the quality of the visual information should produce additive effects with differences due to word frequency and repetition.

Originally, the Logogen System was seen as the interface between all stimulus events which led to verbal, word responses. However, in 1977, Morton reported a series of experiments which showed that repetition facilitates visual word detection if, and only if, the subjects actually see the words in the study phase. Naming words during study in response to pictures of these words or definitions of these words does not confer an advantage in a subsequent detection task using these words. As the Logogen System was developed to explain repetition benefits for <u>same word</u> responses, this failure to find a repetition effect led him to redefine the basic representational level of a logogen. Whereas before a logogen was a single output code, Morton (1977) proposes that each type of input, visual and acoustic, is to be mediated by a separate input Logogen System. The importance that Morton attaches to the effects of repetition in defining the functional level of a logogen's operation is of particular interest. For Morton, any use of a logogen facilitated its later use. However, this was not a prediction within the model, as a change in processing was used to identify the existance of a logogen. As a consequence if performance with a nonword is changed by a prior presentation, then the Logogen System must assume that there are logogens for nonwords as well as words.

<u>The Logogen System and the Lexical Decision Task</u>

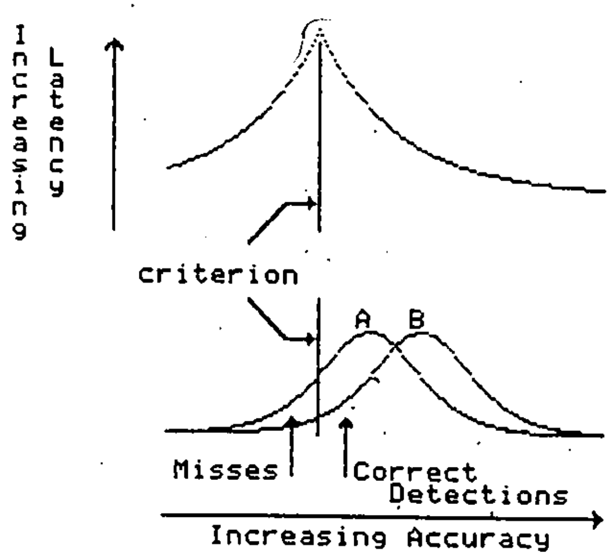
Although the Logogen System was developed originally to describe word detection and not lexical decision, it has often been suggested (Meyer, Schvaneveldt & Ruddy, 1974; Scarborough, Cortese & Scarborough, 1977; Scarborough, Gerard & Cortese, 1979) that performance in the Lexical Decision Task may be described by the Logogen System. Not only do the two tasks require similar information (i.e., they both required word identification), performance in both tasks responded to similar variables, such as word frequency, repetition and semantic context. Interestingly, Scarborough et al. (1979) found an effect of modality in lexical decision similar to that reported by Morton (1977). Like Morton, Scarborough et al. (1979) discovered that there are positive benefits to a prior word presentation if the subject previously saw the word, but not if the subject generated the word as a response to a picture or a definition of the word. Thus, it seems possible that the two tasks share a common resource. Moreover, Morton (Morton & Patterson, 1980) has included the Lexical Decision Task within the realm of phenomena he hopes to explain with the Logogen System. Hence, one objective of this paper is to examine the generalizability of the Logogen System to the Lexical Decision Task.

<u>Dependent Variables</u>. One important difference between the Word Detection Task and the Lexical Decision Task is that the former uses accuracy as its primary dependent variable while response latency is the dependent measure typically used in studies using the Lexical Decision Task. As will be discussed, Signal Detection Theory has no standard method for relating response accuracy and latency. Hence, judging the validity of the Logogen System in lexical decision only by the use of latency data is questionable. The Logogen System, as it is presently stated, is testable only in terms of accuracy. For this reason the studies to be reported will focus on accuracy as well as latency in investigating the Lexical Decision Task.

In light of the difficulty in relating response accuracy and latency most experiments employing a lexical decision attempt to minimize differences of accuracy between conditions by keeping accuracy near a ceiling level of performance. Yet often this is not possible (see Becker & Killion, 1977; Forbach, Stanners, & Hochhaus, 1974; Frederikson & Kroll, 1976; Stanners, Neison, & Painton, 1979). When accuracy cannot be equated across conditions, then an attempt is made to show that accuracy and latency are negatively correlated; and hence that the accuracy data can be interpreted simply as reinforcing conclusions based on latency. It is not the purpose of the present paper to argue with these methods of controlling for accuracy but rather to argue that important information is lost by ignoring the accuracy of lexical decisions. On the basis of data to be reported later, it will be argued that latency differences in lexical decision are constant across large changes in accuracy. Hence accuracy based information is not purchased at the cost of latency based information.

At this point, it will be useful to briefly review the type of difficulty that arises for Detection Theory in describing latency as well as accuracy. Detection Theory (see Atkinson & Juola, 1973) predicts the trade-off typically observed between accuracy and latency by assuming that the mean response time will be inversely and logarithmically related to the distance between the mean of the accuracy distribution and the response criterion. This is diagrammed geometrically in Figure 1. As can be seen, if the events closest to the criterion exhibit the longest decision latencies, the average latency for a signal distribution close to the criterion, distribution 'A', should contain more long latencies than a signal, distribution 'B', which is further from the criterion. As a result, the mean correct decision time expected for

Fig. 1 A possible relationship between **Latency and Accuracy**, based upon Detection Theory.



distribution 'A' will be longer than that expected for 'B'. Indeed, this is a typical result. However, these latency assumptions cannot describe the actual distribution of correct reaction times. Detection Theory predicts that this distribution should have a negative skew (see Figure 2), whereas the empirically observed distribution is postively skewed (Ratcliff, 1978). The expected distribution may be inferred from Figure 1 by noting that a larger number of the responses should have a long latenty, rather than a short latency for correct detection. That is, the largest proportion of the correct responses fall close to the criterion, rather than far from the criterion. As a result, the distribution of correct responses should be negatively skewed, with a few fast latencies and many long latencies. Where the predominant assumptions linking latencies and accuracy in Detection Theory predict that the mean latency will be shorter then the median latency, the actual ordering is the opposite. The mean latency is longer than the median latency. In short, the prediction of response latency based upon response accuracy in Detection Theory is questionable.

<u>Becker's Verifiction Model of Lexical Decision</u>

The Verifiction Model of lexical decisions proposed by Becker (1976), Becker and Killion (1978) is a two process model of word recognition. In this model, the first stage generates a verification set, while the second stage seeks to match the target string with a member of this verification set.

According to Becker (1979) the generation of the verification set is a sensory process. A feature extraction stage provides data in parallel to word detectors, an assumption it shares with Morton's Logogen System. Unlike Morton, Becker proposes that a <u>set</u> of visually similar words is activated during the process of passive data accumulation. It is the job of the second stage, the active verification stage, to select the correct word from this set of

Fig. 2

Alternative Latency Distributions

Expected Distribution: Negative Skew

Observed Distribution: Positive Skew

similar or confusable words. Becker describes the verification stage as an active serial comparison process, similar in nature to serial memory search (Sternberg, 1967). Thus, in the verification stage, the subject's task is to determine whether or not the presented stimulus matches an item in his or her memory set. As memory search is assumed to be serial and self-terminating, then a correct match for items encountered early in the memory search will on the average produce shorter latencies than correct matches for items encountered late in the memory search. Furthermore, Becker proposes that the search stage of set verification is ordered, with frequently seen words usually being verified before infrequently seen or rare words. It is by the use of this verification <u>bias</u> that Becker explains the Word Frequency Effect in lexical decision. In short, frequent words are accepted as words with a shorter latency than are rare words. Because verification is a serial process, the larger the number of words verified and rejected prior to the match, the longer it will take to recognize a word. Hence, low frequency words, being on the average verified after high frequency words, take longer to process than high frequency words.

In Becker's Verification model, the decreased decision latency observed for repeated words is explained in a similar fashion by assuming that a recent encounter with a word will increase the probability that that word will be selected for early verification. In this manner, the benefits of repetition, like those due to word frequency, are attributed to a bias in the order of verification.

While the primary dependent variable in the Verification Model is response latency, response accuracy can be predicted with an additional assumption (O'Conner & Forster, 1981). Consider the case where the verification process is stopped after a set period of time. As verification is ordered by frequency, this truncation would result in a greater number of low frequency words failing to be verified than high frequency words. When a process involved in serial search stops randomly before all items are searched then not only are entries at the end of the list searched later than initial entries, there is also a greater probability that they are not tested at all. A similar comparison can be made for the effect of repetition.

<u>The Logogen System and the Verification Model Compared</u>

As the Logogen System was designed to explain decision accuracy and the Verification Model was designed primarily to account for decision latency, it is impossible to directly contrast the two theories without making specific assumptions about the relationship between the accuracy and the latency of responding. Rather than seeking to generate predictions that will allow us to choose between the Logogen System and the Verification Model as models of the Lexical Decision Task, the empirical evidence discussed below demonstrates that even within their own response measures of the Lexical Decision Task, neither model offers a sufficient explanation. However, it will be found that both the accuracy and latency data are compatible with predictions based upon a model of Relative Familiarity Judgment.

<u>Relative Familiarity Judgments</u>

The Relative Familiarity Judgment model addresses two issues. First, it is assumed that subjects do not rely on a single resource in making a lexical decision; rather, it is assumed that the Lexical Decision Task is a relative and multi-dimensional judgment, drawing upon one or several discriminative dimensions as the situation demands. Secondly, it is assumed that the latency of responding is linked to the accuracy of responding and is described by the Random Walk Diffusion Process.

The Relative Familiarity Judgment model assumes that there is not a specific collection of detectors in memory which guarantee that a word is present. Rather, it assumes that there are only memories about things that we have experienced and call words, memories about what they look like, what they imply and how they can be used as semantic tools. A word is a word because we know something about it. A string is rejected as a word if we know very little about it, if we cannot remember having seen it, and if it does not suggest a semantic context. To the extent that a nonword fails to be distinct from past experience, it is likely to be called a word. This simply recasts the problem by emphasizing memory for specific attributes rather than memory for regularities. Yet it is this emphasis on attributes, that in part supports the relativity of a word judgment in the relative familiarity model. A word is not classified by reference to a specific memory or memory systems; rather there are converging lines of evidence, and it is suggested that subjects respond to this confluence as well as to the specific attributes.

Other factors that lead to a relative judgment are the encoding and decision process. In the relative familiarity model, the subject's task is to discriminate words from pseudowords on the basis of the encoding activity. As pseudowords are designed to be confusable with words, then it is to be expected that pseudowords as well as words will lead to significant encoding activity. The relative familiarity model assumes that subjects, in order to discriminate between word and pseudoword encoding activity, compare the ongoing encoding activity relative to a standard or expected level of activity. Word decisions are given for those stimuli which over time yield richer than normal encodings, while pseudoword decisions are given for those stimuli which over time yield weaker than normal encodings.

In the Relative Familiarity Judgment model, word decisions for frequent stimuli such as house or theory, which have multiple and rich representations both visually and semantically, are highly probable at any level of comparison. In opposition, with less well known stimuli such as adobe, skirl, and stune, word responses are less probable and are more likely to depend upon task demands. Stune, while visually familiar (It was used earlier as an example of a pseudoword) is not a word. However, it is possible that this episodic increase in familiarity might lead to an incorrect word categorization for this familiar pseudoword. This is especially probable if decisions are based primarily upon visual familiarity, such as when the nonword distractors are visually and structurally unfamiliar (e.g., gozpb). At the same time, word decisions are likely to be given for adobe and skirl. In contrast, when the pseudoword distractors are visually and structurally familiar and hence hard to reject on the basis of visual familiarity, then we would expect that a word decision would rely more heavily on meaning. In the latter case, decisions would be less sensitive to the increased familiarity of a familiar pseudoword, and at the same time less accurate for adobe and skirl. In fact, decisions might be delayed until a context could be generated for these words. Moreover, we might expect differences in the decisions to adobe and skirl because adobe has a readily specified referent (i.e., brick-like) whereas skirl is associated with a context of use (i.e., bagpipes) rather than an imageable referent. In short, it is proposed that people make lexical decisions on a relative basis, not on an absolute basis. It is relative because the selection of relevant information will depend upon both decision difficulty and task demands. To this extent, then, lexical decision, and by implication word recognition, cannot be described as depending upon automatic and context free detection mechanisms. While the contact of the current encoding operation with similar events in memory may be automatic, this contact does not constitute by itself a word decision or a word detection since the usefulness of this contact depends on the task demands. Familiarity, as it is currently used, simply reflects the degree or the extent of this contact with memory and is assumed to be an automatic consequence of a test encoding, although again not a word decision in itself.

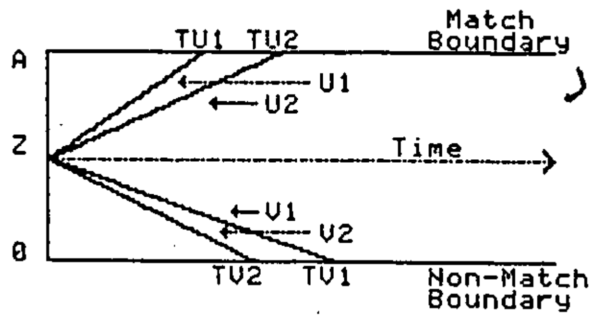
The Relative Familiarity Judgment model differs from conventional lexical store models of lexical decision primarily in the predictions it makes about the processing of pseudowords and in the inclusion of a task relative decision criterion. The Lexical Decision Task is assumed to require a relative judgment because a change in the decision criterion is assumed to correspond to a functional change in the type of information, visual or semantic, that is used in making the decision. The lexical store models assume that the lexical decision process is essentially insensitive to nonwords as visual stimuli, and that Lexical Decisions are confined to a single information base, while the relative familiarity hypothesis predicts that decision performance will be sensitive to nonword familiarity and to multiple word features.

<u>The Random Walk Diffusion Process</u>. The quantitative basis of the current Relative Judgment model of lexical decision is based upon work published by Link (1975, 1978), Ratcliff and Murdock (1976) and Ratcliff (1978) using the Random Walk Diffusion Process as a model of memory based decision. The random walk process was used by both these researchers to model the interactions of memory and perceptual encoding leading to a memory decision. In the present discussion, the implications of this framework for memory decisions will be explored. There are three benefits to thinking of the Lexical Decision Task in

terms of a random walk model. First, unlike Signal Detection it supplies a framework which explicitly addresses both accuracy and latency of responding; second, it provides a concrete metaphor for the often vague concept of familiarity; and third, it yields a useful organization of the costs and benefits associated with familiarity.

There are three major variables in the Random Walk Model which are relevant to decision performance. These are 1) the overall level of caution, A in Figure 3, 2) the starting value for the comparison process or response bias, Z in Figure 3, and 3) the similarity of the test target to the memory referent, represented by the slopes $U1$ and $U2$ in Figure 3. For any given decision, it is assumed that the decision process begins at Z, which for the moment we will assume is equal to one half A. The model assumes that the features of a stimulus are encoded in a time dependent comparison process by contrasting the featural information with information in memory. For the present, assume that the referent is discrete for each item. The comparison process accumulates a weighted sum of these feature comparisons, in which a feature match will yield a count of +1, and a non-match will yield a count of −1. For each stimulus, the comparison process begins at Z which represents the a priori expectation of receiving either the target or distractor. The slopes, U and V, represent the cumulative sum of the comparison process over time added to Z, where U is a target and V is a distractor. Whenever the sum of the comparisons equals A, the decision process signals a match; and whenever the sum of the comparisons equals zero, the decision process signals a non-match. For any given target there is a certain, albeit often small, probability that the target may cross either the match or the non-match barrier. The probability and the latency with which a target will cross the match barrier rather than the

Fig. 3 Geometric Analog of The Random Walk Diffusion Process.



Where: A > Z > 0 : and
U and V are Slopes based on Similarity

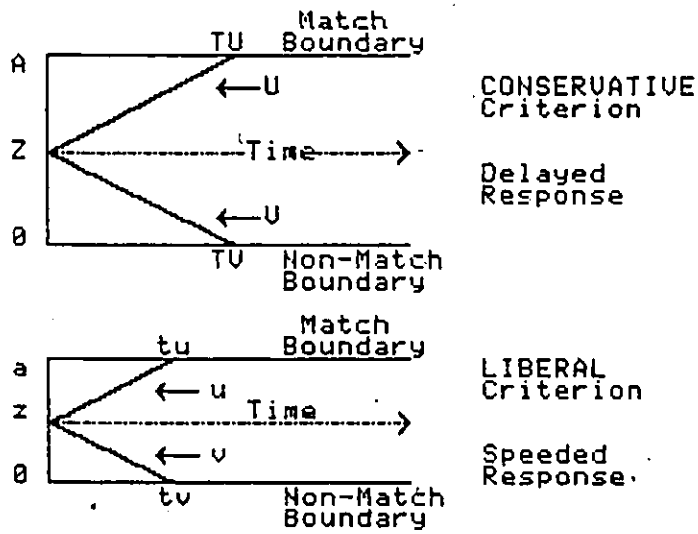
non-match barrier increases as the similarity of the encoded target relative to the comparison referent in memory increases. In a similar manner, the probability and latency of rejecting a distractor are tied to the similarity between the target and the memory based referents.

A well known phenomena in studies of discrimination and detection is the reciprocal relationship between speed and accuracy (Posner, Klein, Summers and Buggie, 1973). This relationship is captured in the Relative Judgment Model by varying the level of caution, variable A in Figure 3, the amount of evidence required to make a response. If A is relatively large in value, then accuracy will increase at a cost of increased response latency. In contrast, if A is relatively small in value, then decision latency will be speeded at a cost of decreased accuracy. This is shown visually in Figure 4.

Response bias in the Random Walk Model is associated with variable Z. If Z equals one half A, there is no response bias. If Z is less than one half A, there is a response bias for responding non-match. If Z is greater than one half A, there is a response bias for responding match. The consequence of a response bias, a match bias, for example, is to decrease the information necessary to make any match response while increasing the information necessary for a non-match response. A non-match bias reverses the demands on information.

An important similarity between response bias and a speed accuracy trade-off is that neither manipulation affects the slopes of the diffusion process. The difference between response bias and speed accuracy trade-off is that in a response bias the latency and the errors decrease at one decision barrier while they increase at the other barrier, while there is an overall increase or decrease in latency and accuracy in a speed accuracy trade-off.

Fig. 4 Speed Accuracy TRADE-OFF in the RANDOM Walk Process.



Where: $a < A$

and: U, u, V, v are slopes.

As the value of parameter (A, a) increases the level of ACCURACY will also increase.

Further, as (A, a) increases the intercepts for the slopes U, u, V, v will also increase.

Memory Retrieval in Relative Familiarity Judgment. Like the Logogen System, the relative familiarity model does not use a process of serial look up to explain the latency and accuracy of memory decisions, but instead focuses upon decision difficulty. Specifically, Relative Familiarity Judgment assumes that the extraction of the similarities and the discrepancies between a given target stimulus and a standard in memory account for item differences. Nonetheless, it must postulate some process in which the stimulus can interact with memory. Ratcliff (1978) accomplishes this by using the resonance metaphor. The resonance metaphor is based upon the analogy of acoustic resonance. In this analogy, probe items are likened to a vibrating tuning fork with an unknown frequency, while memory is likened to an array of static pitch forks, all of differing frequency. Thus, as in acoustic resonance, because of structural considerations, the probe item will activate only structurally similar items in memory. Such a process, one of parallel comparison, is similar to a retrieval system in Gestalt Psychology and Pandemonium (Selfridge, 1959) wherein memory contact is a direct function of the similarity between ongoing perceptual processes and the structure of the memory traces. The critical feature of these systems is that this parallel process eliminates a sequential search of memory for the target, and instead emphasizes the match process itself.

Relative Judgment Theory (Link, 1975) assumes a single standard of comparison for memory judgments. In the typical experiment, subjects are given the standard for comparison. However, Link (1981) agrees that the memory standard in the Lexical Decision Task must be implicit and not explicit. Dr. Link proposed that the standard for comparison in lexical decision can be considered to be a particular area in a multidimensional memory space, defined by its similarity to the stimulus. This area will correspond to the psychological representation of our experience with words, although generally, only visual characteristics will form the functional memory standard in lexical decision. Hence, the memory standard is a convenient abstraction and actually refers to some form of psychological representation of past encodings. However, for any particular stimulus, the memory standard will reflect only those memory traces that share features in common with the current encoding of the target stimulus. This is a form of content addressable memory. Memory retrieval in this description is passive and is similar to the resonance metaphor in its appeal to intrinsic structure. As an aside, this perspective of the memory standard is similar to process descriptions of human memory (Kolers & Smythe, 1979), and to instance models of concept formation (Vokey & Brooks, 1982). A more detailed examination of this similarity may prove to be useful. However, regardless of the details of the memory representation, Link assumes that lexical decision reflects a comparison process, which contrasts the current stimulus encoding with past encodings of structurally similar stimuli.

In conclusion, both Ratcliff and Link hypothesize similar processes of memory retrieval in which the current encoding of the stimulus automatically accesses similar encodings in memory. Moreover, for both Ratcliff and Link, this automatic contact between the current encoding and similar past encodings defines the memory referent or memory standard. While past events are necessary in establishing the memory referent, not all familiar past events are equally relevant. Thus, while the number "25" may be highly familar, it is not likely to be confusable with a word. Clearly, there must be a task dependent restriction of prior memories that enter into the memory standard. This cognitive component will be discussed more fully in the following discussion of the relatedness dimension.

The Relatedness Dimension

It has been suggested that the origin of processing differences for targets varying in familiarity are due to differences in the similarity between the test item and a memory referent. A useful method for conceptualizing this similarity is Ratcliff's relatedness dimension. Ratcliff defines relatedness as the ratio of the total number of features shared by a memory referent and its target over the total number of features encoded for the target. This relationship is shown in Figures 5a and 5b. Distribution U_1 in Figure 5a represents a well known target, with a relatedness of .8, while distribution U_2 in Figure 5b represents a less well known target, with a relatedness of .5. When applied to the Random Walk Diffusion Process, these relatedness values produce differences in slopes, similar to U_1 and U_2 in Figure 3.

Figures 5a and 5b contain two additional distributions labeled V_1 and V_2. These are hypothetical, novel distractors, each of which bears a 50% similarity in the total number of features shared with targets U_1 and U_2. As a consequence of this similarity, the novel distractors will have a non-zero value on the relatedness dimension. This value is .4 for V_1 (.5 * .8), and .25 for V_2 (.5 * .5). Notice, though, that while V_1 and V_2 have the same formal similarity with U_1 and U_2 (set arbitrarily at 50% here), the distractors should not be equally confusable with the targets in a memory based comparison. The reason for this is that the distance between the target and distractor corresponds to a difference in d' in Signal Detection Theory. As the distance between the target and distractor is larger in Figure 5a than in Figure 5b, the prediction that follows is that it should be easier to discriminate V_1 from U_1 than to discriminate V_2 from U_2. In short, increasing relatedness implies increasing

Fig. 5a High Relatedness Targets: with 50% Similar Distractors.

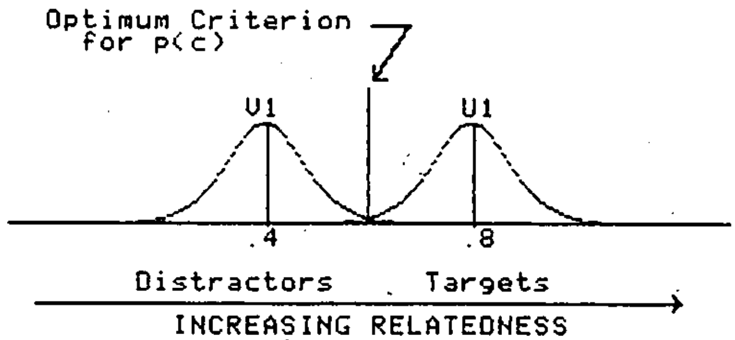
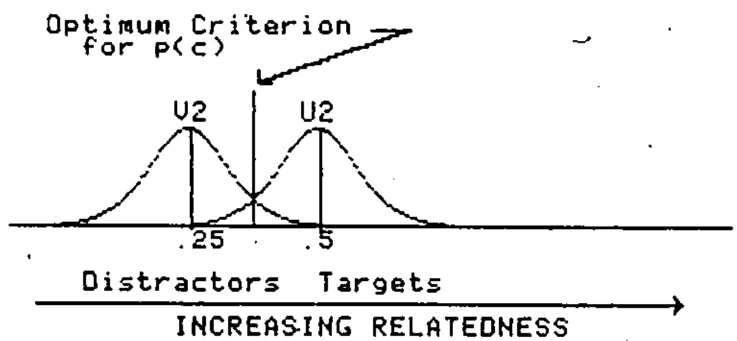


Fig. 5b Low Relatedness Targets: with 50% Similar Distractors.



discriminability from featurally similar distractors. This, of course, is in agreement with common sense, and is only a fancy way of saying that well known items are highly discriminable.

It should be noted that the relatedness criterion and the relatedness dimension are only indirectly associated with the random walk process, and are used only to describe the inferred memory structures that lead to item differences in the random walk process. Moreover, response bias (Z) and the relatedness criterion are independent factors; Z is an priori estimate of which response will be required, whereas the relatedness criterion is an a priori estimate of the average test stimuli's familiarity. A change in the response bias affects the starting value but not the rate of change in the comparison process. A change in the relatedness criterion affects the rate of change in the comparison process and may even reverse the standard response given to some stimuli. Thus, where a change in the relatedness criterion will affect the degree of slope for U and V, a change in response bias (Z) will not affect the slopes.

The effective relatedness dimension in any given experiment is formed in the interaction of the task demands (e.g., type of stimuli and instructions) with the personal experiences of the subjects. The formation of the relatedness dimension is, in part, context sensitive. Hence, the information used for a lexical decision is modified by the task demands, and to a limited degree can be modified to maximize specific goals. The relatedness dimension appealed to is in one sense real, as it is assumed to reflect objective relations between the encoding of the target and memories for similar prior instances. However, in another sense, the relatedness dimension is arbitrary in that it forces stimuli which may vary across multiple dimensions onto a common continuum. While this assumption of a single relatedness continuum may eventually require modification, it will serve as a useful first approximation.

To return to Figure 3, U_1 and U_2 are shown as having the same origin, but different slopes. The origin represents the initial evidence available for responding. At the start of the comparison process this evidence is assumed to be similar for all items. In processing, item differences will be expressed as a difference in the ratio of feature matches to feature non-matches (the difference ratio). Familiar items will possess a large difference ratio (feature matches will outnumber feature non-matches) and the value accumulated by the comparison process will rapidly approach that necessary for a positive response. Thus, U_1 in Figure 3 is drawn with a steep slope to show a rapid accumulation of evidence. Unfamiliar items will produce a smaller difference ratio, and consequently the comparison process will approach the decision barrier at a slower rate. Therefore U_2 in Figure 3 is drawn with a shallow slope to depict a slow accumulation of evidence. Hence, as it is assumed that the proportion of feature matches to feature non-matches is larger in U_1 than in U_2, then the rate and reliability with which stimulus information is accumulated, and hence, the rate and reliability with which the sum of $\underline{Z}$ and the stimulus evidence appoaches and equals $\underline{A}$, will be greater for U_1 than for U_2.

In conclusion, the relative familiarity model assumes that item differences in decisions arise from differences in the stored information about target items; and this memory difference results in a processing difference. This processing difference leads to item differences in decision accuracy and latency. Hence, it is assumed that in a Lexical Decision Task the similarity of the present encoding to similar encodings in memory will govern the rate with which information is accumuluated for any response. On average, well known targets will lead to a faster and more reliable of accumulation of evidence than will poorly known targets.

The Nature of Familiarity

Familiarity is a common catch word in psychological research that often suffers from a surfeit of familiarity. It is overused, and often fuzzy in definition. Nonetheless, it is too useful a term to banish simply on these grounds. For that reason, in the present paper familiarity will be used as a major psychological variable. However, in this paper familiarity will be synonymous with Ratcliff's relatedness measure. Hence, familiarity in the present context will appeal to a measurable relationship between the test stimuli and memory.

Familiar items are not only common, they are also well known. Thus, familiar words are not considered to be specially marked or tagged words in a detection system or look-up table, as in the lexical store models. Rather, familiar words are frequently used words, words for which subjects have rich internal representations, and words for which subjects have available multiple skills. In short, familiarity is assumed to be a descriptive statistic. Consequently, when it is claimed that decisions are based on familiarity, familiarity is not seen as a value extracted and assessed by the subject; rather familiarity simply refers to the objective relationship between the target and memory for previous experience with similar targets, which has the consequence of producing detectable differences in stimulus encoding.

One consequence of increased familiarity within the Relative Judgment Model is a steeper drift towards the positive or match barrier. This results in an increased probability and a decreased latency of responding yes for familiar targets, when compared to less familiar targets. The Relative Familiarity Judgment Model of lexical decision accounts for the effects of word frequency and repetition by the uses of familiarity. Put simply, the likelihood of an adequate match, defined by the level of caution parameter <u>A</u>, in a given time period is higher for High frequency and repeated words than for rare words, as the value of the drift parameter (U_1) for frequent words is greater than the drift parameter (U_1) for rare words. As will be recalled, the value of the drift parameter 'U' is directly linked to the difference ratio, which in turn is set by the level of relatedness.

<u>Relative Familiarity</u>. In the previous discussion of similarity and distractors in the Lexical Decision Task, it was proposed that the relatedness criterion used in an easy lexical decision would likely differ from that used in difficult decision. One basis of this assumption is shown in Figures 5a and 5b. Assuming that the relatedness distributions are similar to signal detection curves, then it can be seen that the optimal relatedness criterion for distinguishing between U_1 and V_1 (.60) is not the optimal criterion for distinguishing between U_2 and V_2 (.45). Indeed, as Figures 5a and 5b represent two different discrimination experiments, the two relatedness criteria will be unrelated, and should depend only upon the distributions of the two stimuli to be discriminated. This independence in the selection of the relatedness criterion is one reason for asserting that the judgments of familiarity are relative.

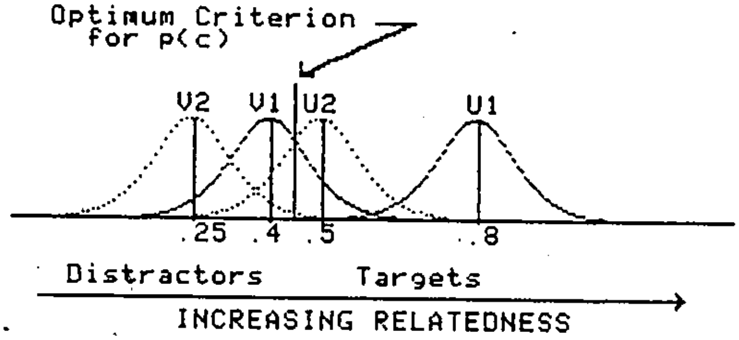
<u>The Word Frequency Effect Considered as a Passive Encoding Bias</u>

An important finding for studies of word identification is that high frequency words are not only more likely to be detected than low frequency words, they are also more likely to occur as incorrect responses. That is, incorrect word responses, or false alarms, are more likely to be frequent words than rare words (Broadbent, 1967; Broadbent & Broadbent, 1975; Catlin, 1973). Moreover, if the hits and false alarms are analyzed in terms of Signal Detection theory, by computing a normalized accuracy score, the increase in hits appear to be largely offset by the increase in false alarms. Broadbent (1967) and Broadbent et al. (1975) argued that this trade off between hits and false alarms with word frequency was compatible with a stimulus contingent response bias, but not a stimulus contingent encoding bias. A stimulus response bias differs from a simple response bias in that the bias in responding is contingent upon the a priori occurrence of a word given certain stimulus information, and is not simply due to the overall a priori probability of a word's occurrence. A stimulus response bias is contingent upon encoded featural information; and the effective value of the bias can be considered as changing with the type of stimulus information available. An encoding bias differs from a response bias in that an encoding bias results in the preferential encoding of information relevant to the stimulus. Thus, the encoding bias hypothesis predicts that more information will be available about features characteristic of high frequency words than features charactersitic of low frequency words. Since more information implies greater distinctiveness, the encoding bias hypothesis can be seen to predict greater precision for judgments about frequent words than rare words. In the terms of Signal Detection Theory estimated d' for frequent words should be larger than the estimated d' for rare words. The stimulus response bias hypothesis predicts no difference in d' for frequent and rare words, but a difference in Beta. As mentioned earlier, the experimental evidence found that as word frequency increased the proportion of frequent words given as false alarms also increased. Within detection theory, this increase in false alarms comes from a change in the response criteria (Beta) rather than a change in response precision (d'). Hence, Broadbent argued that frequency in the language is more compatible with a stimulus response bias than an encoding bias.

In the Lexical Decision Task, as in Word Detection Tasks, increased accuracy for high frequency words generally is accompanied by the presence of increased errors for pseudowords similar to high frequency words (O'Connor and Forster, 1981). Thus, as in word detection, frequency effects in lexical decision would appear to be due to some form of response bias, rather than an encoding bias. However, in the earlier discussion of relatedness, it was claimed that relatedness differences resulted in encoding differences. With increasing relatedness there should be an encoding bias, which implies a difference in d' and not Beta. The data reported by O'Conner and Forster (1981) require a difference in Beta with word frequency. Hence, it appears that the data do not agree with the predictions made by the Relative Familiarity Judgment model.

A detailed examination of judgments based on a relatedness continuum, however, reveals that in the typical lexical decision experiment relative familiarity predicts an increase in both d' and Beta with increasing familiarity. That is, there should be more false alarms for distractors similar to familiar targets than for distractors similar to unfamiliar targets, although the normalized distance between targets and distractors should be larger for familiar targets than for unfamiliar targets. The basis for this prediction is the use of multiple targets and distractors with a wide range of familiarity in the typical lexical decision experiment. This relationship is presented graphically in Figure 6. The distributions shown in Figure 6 for targets U_1 and

Fig. 6 Variable Relatedness Targets with 50% Similar Distractors.



U_u and distractors V_f and V_u are identical to those diagrammed in Figures 5a and 5b and have the same relatedness values. In Figure 6, however, they have been placed upon the same decision axis, representing an experiment which uses a mixed list of familiar and unfamiliar targets and their distractors. Notice that the distance between each target distribution and the distribution of its distractor is the same as in Figures 5a and 5b. Hence, <u>d'</u> for familiar stimuli (U_f–V_f) is greater than <u>d'</u> for unfamiliar stimuli (U_u–V_u). Notice, however, that there is a new decision criterion. This criterion is set to optimize the overall probability correct for decisions between the two target (U_f, U_u) and the two distractor distributions (V_f, V_u); because of the difference in overall familiarity between the two sets of distractors, the distractors from V_f are expected to produce more false alaams than the distractors from V_u.

As can be seen, the Relative Familiarity Judgment model predicts that with increasing discriminability there can be a correlated increase in the false alarms to similar distractors. This is the result reported by O'Connor and Forster (1981). A reanalysis of their results revealed that despite the greater false alarms for distactors similar to familiar words, the estimated <u>d'</u> for discriminations between familiar words and their distractors was larger than the estimated <u>d'</u> between unfamiliar words and their distractors. Hence, the relative familiarity hypothesis predicts the pattern of errors found by O'Connor and Forster (1981) with word frequency and related non-word distractors in the Lexical Decision Task. An explanation of the effects of word frequency in lexical decision in terms of a stimulus response bias , such as that used in the Logogen System, requires additional assumptions to account for the correlated increase of <u>d'</u> and <u>Beta</u> observed with increasing familiarity.

The preceding analysis reveals an interesting relationship among signal detection curves. In Signal Detection Theory, as described by Green and Swets (1964), comparisons are typically discussed in terms of discriminations between a single signal and a single signal-plus-noise distribution, in which case the relationship between d' and Beta will be orthogonal. That is, a change in one descriptor cannot predict a change in the other descriptor. However, in cases where a single response criterion is used with three or more distributions, d' and Beta can be linked. Thus, as in Figure 6, if two or more target distributions, that differ in their means, are compared with the distribution of their distractors then estimates of Beta will increase as estimates of d' increase. This has important implications for psychological research, given the usual variety of different stimuli tested in an experiment, as it implies that the origin of Beta effects may be due to the same factors that give rise to d' differences. In other words, differences in the estimated Beta between several distributions may be an artifact of the comparison process and a biased set of memories, rather than an independent psychological entity. An encoding bias imparted by increased familiarity may result in dissimilar d's and similar Betas as in Figures 5a and 5b, or it may result in dissimilar d's and dissimilar Betas as in Figure 6. The actual relationship between d' and Beta will depend upon the task and the type of stimuli tested. In conclusion, it is neither necessary nor, perhaps, appropriate to invoke separate psychological mechanisms to explain d' and Beta effects.

The Relevance of the Theoretical Discussion for the Following Experiments

The following will list the assumptions shared by the Logogen System and the Verification Model as examples of the lexical store model of the Lexical Decision Task. A following section presents the general assumptions made by the relative familiarity hypothesis.

The lexical store explanations of the Lexical Decision Task make the following assumptions: 1) The first stage in any use of a word is making contact with a specialized lexical memory. 2) The Lexical Decision Task is assumed to be a direct measure of this first stage. 3) Only meaningful words are represented in this lexical memory. 4) The increase in speed and/or accuracy of lexical access produced by a prior presentation of a word is independent of the subsequent use of that word. 5) The information in lexical memory, and hence the information used during access, is independent of the specific form of words (i.e., variations in case and typography), although lexical memory might be modality specific. 6) The detection of a word unit is all or none and partial information is unavailable for lexical decision. 7) The Lexical Decision Task is a direct and unbiased measure of a skill used in reading. Issues three to six will be addressed directly in the following experiments, while issues one, two, and seven will be only indirectly addressed.

I will argue that a lexical decision may draw upon a broad spectrum of episodically based information and that, potentially, any previous encounter with a stimulus, or with a similar stimulus can provide information for a lexical decision. Hence, a pseudoword, if for some reason it has been seen before, may some times be accepted as a word. In short, lexical decisions are not restricted to words (i.e., legitimate vocabulary) but rather are based upon a salient similarity between the target and visually based memory information about graphemes. While there is selectivity, the task instructions and the test stimuli provide general constraints for a decision, but lexical status must be inferred not looked up. In analogy with tests of recognition memory, there are multiple ways of accomplishing a lexical decision, none of which is always correct. If pseudowords are highly similar to, and hence confusable with, words, then selection at a semantic level will accept these similar pseudowords. If pseudowords have been previously seen, then selection at a visual level may accept these familiar pseudowords. Furthermore, as the focus of a decision shifts between stimulus properties, the relative importance of variation in other properties becomes less important. If a decision focuses on visual properties, variation in semantic content will be trivialized. Likewise, if a decision focuses on semantic content, then that decision will likely be unaffected by variation in visual characteristics, such as typography and case. Therefore, as in recognition memory, the relevant information supporting a lexical decision will depend upon the type of the stimuli used as targets and distractors. Such considerations as the familiarity of the stimulus, for words and pseudowords and the availability of meaning for words will exert different effects in different situations.

In conclusion, the hypothesis to be investigated in the following experiments are that 1) lexical status is inferred on the basis of multiple sources of information, 2) lexical decisions are based upon a flexible use of memory information and respond to task demands, and 3) the inference of lexical status can be drawn from information contained in a single instance as well as information known about words.

Experimental Data

The following five experiments were designed to investigate the types of knowledge subjects typically employ in the Lexical Decision Task. The experiments employed manipulations designed to modify a subject's task strategy by 1) changing the visual properties of a grapheme and hence in one sense the familiarity of that grapheme, and 2) selecting pre-experimental properties of the stimuli, such as word frequency and the type of meaning associated with a word, e.g., abstract or concrete reference, or orthographic structure with pseudowords. Following each experiment, there is a brief discussion of the theoretical implications of the results in terms of the three models of the Lexical Decision Task discussed in the introduction.

Experiment 1 was designed 1) to test for multiple bases of information in the Lexical Decision Task by investigating the importance of phonemic, semantic, and visual information; and 2) to provide a test of whether lexical decisions can be based upon semantically neutral and episode-based visual information. Experiment 2 was designed 1) to provide additional evidence for the use of semantic information in the Lexical Decision Task by experimentally manipulating an episode-based availability of semantic information, 2) to investigate the importance of task demands in determining the type of information used in making a lexical decision, and 3) to assess the importance of partial information in lexical decisions. Experiment 3 was designed to contrast the contribution of specific visual information against letter-based information in the repetition effect. Experiment 4 was designed to investigate the possibility that the effects seen for frequency in the language and for semantic information were based upon similar memory information. Experiment 5 differed from the first four experiments in that it focused exclusively upon the quantiative predictions of the models of lexical decision, and was designed to link the additive effects of word frequency and task demands observed in Exp. 2 to the additive effects reported for word frequency and visual quality in the literature on lexical decisions (Stanners, Jastrzembski & Westbrook, 1975; Becker & Killion, 1977).

Experiment 1

It was suggested earlier that phonemic information, visual information, and meaning might serve as separate sources of information that are utilized when subjects make lexical decisions. In determing whether or not phonemic information is employed has proven to be difficult. While some studies have found evidence of phonemic information (Meyer, Schvaneveldt & Ruddy, 1974; Rubenstein, Lewis & Rubenstein, 1971), other studies have indicated that phonemic information is not necessary (Kleiman, 1975; Shulman, Hornak & Sanders, 1979). Lexical decisions are slowed when nonwords are pronounceable Balota and Neely (1981). This observation could be taken as evidence that phonemic information is involved and that subjects pronounce test words as a means of gaining access to meaning or perhaps even make their lexical decisions on the basis of the ease of pronounceability. However, pronounceability is necessarily highly correlated with the visual characteristics of a letter string (Frederikson & Kroll, 1976). It may be that these visual characteristics are responsible for the effects observed by Balota and Neely (1981). In the present experiment, subjects in one condition were required to pronounce test items as well-as make a decision about the lexical status of the test items, while subjects in a second condition only made lexical decisions. Adding the requirement to pronounce test items provided a means of assessing the extent to which phonemic information is typically utilized when subjects are asked to make lexical decisions. If phonemic information is typically utilized, the requirement to pronounce test items should have little effect since even without this requirement subjects must typically pronounce those items as a means of gaining access to phonemic information. In contrast, if phonemic information does not typically serve as a basis for lexical decisions, the requirement to pronounce test items has the status of a secondary task; effects on ease of pronunciation should be separate from effects on lexical decisions. The parallel manipulation of pronunciation and lexical decision should interact with several factors including frequency in the language and, perhaps, repetition (see Becker, 1976). High frequency words are pronounced more rapidly than are low frequency words (Frederikson and Kroll, 1976; Scarborough et al., 1977) so the requirement to pronounce should add less to the total reaction time of high frequency than to that of low frequency words. With regard to repetition, pronunciation of a word might benefit from repetition so that pronouncing a repeated word adds less of a load than would pronouncing a word on its first presentation.

Another test of phonemic information involved the manipulation of syllabic structure. The time to onset of naming has been found to increase as the number of syllables in a word increase (Eriksen, Pollack & Montague, 1970; Frederikson, 1979). This is often especially noticeable in pseudowords. A possible interpretation of this is that pronouciation for multisyllabic pseudowords involves an extra process of vocally combining the code for single syllables, and that naming with pseudowords requires pronouciation by constituents more than naming with words. By examining the effects of syllabic structure when lexical decisions were made alone and with a secondary pronouciation task, it was possible to assess the effect of pronouncability in a typical lexical decision.

Effects of repetition were also of interest. Prior experiments have revealed that repetition interacts with a number of factors. The large effect of frequency in the language that is found for the first presentation of a word is greatly diminished when words are repeated (Forbach, Stanners &

Hochhaus, 1974; Scarborough, Cortese & Scarborough, 1977). This observation can be taken as evidence that the effects of frequency of presentation are similar to those of frequency in the language. Of greater interest for the present experiment, variation in the meaning of test words has also been found to interact with repetition. James (1975) found faster reaction times to words with a concrete referent than to those with an abstract referent. This difference in reaction time was greatly reduced when words were repeated. A possible interpetation of this result is that subjects rely on meaning less when making a lexical decision for repeated words, and instead rely more heavily on the visual characteristics of the word. In the first experiment, an attempt was made to replicate the interaction between concreteness and repetition. Later experiments investigated further the role of meaning and the role of visual characteristics as bases for lexical decisions.

Also of interest was the nature and origin of memory information used in making a lexical decision. McKoon and Ratcliff (1979), Balota and Neely (1980) and Carroll and Kirsner (1982) found that pseudowords previously encountered in a paired associates study list showed longer rejection times and more errors than novel pseudowords. This suggests that the memory information used in lexical decisions can be an episodic trace as well as a generalized lexical store. In the first experiment an attempt was made to replicate this result.

Method

Subjects. Thirty two McMaster undergraduates received an hour's course credit for participating in the experiment. At the start of a session a subject was randomly assigned to one of two groups, an unbiased or a pronunciation-biased group, with the provision that 16 of 32 subjects served in each group.

<u>Materials</u>. Two sets of 40 words (six letters in length) were selected from the Thorndike and Lorge (1944) norms, with the provision that all words in one set were High Frequency Words (<u>A</u> class words, with frequencies between 50 to 100 times per million), while words in the other set contained Low Frequency Words (with frequencies of 1 to 2 times per million). Moreover, selection was constrained so that half the words in each set were judged to have concrete referents while the other half had abstract referents. A concrete word was defined to be any word whose referent was a sight, sound, or touch that was readily available to the senses and could be directly experienced; abstract meanings were defined to be any non-sensible event or relation. Low frequency, abstract words had slightly higher mean frequencies than the rare concrete words (1.65 vs 1.60). A further constraint used for word selection was that half the words in each frequency set, and half the concrete and half the abstract words, were monosyllabic words, while the other half were bisyllabic words.

An equal number of six letter pseudowords were generated under the constraint that all pseudowords were orthographically legal and pronounceable. In addition, half the pseudowords were monosyllabic and half the pseudowords were bisyllabic. No pseudoword was a homophone for an english word.

An additional 30 words and 30 pseudowords were collected for a practice session, following the above procedures, for words and pseudowords.

<u>Design</u>. An experimental session consisted of two phases, a study and test phase. The study phase (Phase I) was a Lexical Decision Task, and served to experimentally pre-expose half the words and half the pseudowords. The experimental phase (Phase II) was also a Lexical Decision Task and used all 160 test stimuli. The stimuli were counterbalanced across subjects such that

half the stimuli were seen by half the subjects in both Phase I and Phase II, repeated stimuli; and half the stimuli were only seen in Phase II, experimentally novel stimuli. Old and new stimuli were conterbalanced between subjects.

The between groups manipulation of verbal coding was produced by varying instructions. Subjects in the unbiased group were instructed to respond at the onset of the target stimulus as quickly and as accurately as possible with a lexical decision. The pronunciation-biased group received the same instructions but, in addition, were required to name the word or pseudoword seen, <u>after</u> performing the lexical decision. Subjects receiving the secondary task experienced little difficulty with the secondary task and named the stimuli in the practice trials, in Phase I, and in Phase II. The pronunciation-biased group was told that their verbal responses were being recorded, and would be scored for accuracy at the end of the experiment. No such data were gathered.

<u>Procedure</u>. All subjects were positioned two feet away from a video screen. A millisecond timer was activated concurrent with the onset of a stimulus on the screen, and was terminated by a key press. The stimulus was erased from the screen one half second after the key press. Approximately three seconds elapsed before a warning signal, consisting of three plus signs was shown for 750 msec, which in turn was followed 500 msec later by the onset of another stimulus.

Stimuli were presented by a Digital PDP-8 computer on a green phosphor video terminal. Binary responses were collected by the computer, following a button press from one of two response buttons, to the nearest millisecond.

TABLE 1

Reaction times (in msec) and Percent Accuracy for Words, in Experiment 1.

CONTROL GROUP

	HIGH FREQUENCY WORDS		LOW FREQUENCY WORDS	
MEANING:	CONCRETE	ABSTRACT	CONCRETE	ABSTRACT
	RT ACC.	RT ACC.	RT ACC.	RT ACC.
NEW	542 (98.8)	569 (95.6)	601 (85.0)	628 (81.9)
OLD	518 (100.)	529 (98.8)	558 (96.9)	551 (99.4)

EXPERIMENTAL GROUP

	HIGH FREQUENCY WORDS		LOW FREQUENCY WORDS	
MEANING:	CONCRETE	ABSTRACT	CONCRETE	ABSTRACT
	RT ACC.	RT ACC.	RT ACC.	RT ACC.
NEW	544 (97.5)	591 (95.6)	649 (91.3)	693 (87.5)
OLD	522 (100.)	530 (98.8)	578 (93.8)	612 (94.4)

Results and Discussion

The mean latency and accuracy for each condition in Phase II was computed for each subject. These means were entered into an analysis of variance. No significant or near significant effects were found involving syllable length of words, in either latency or accuracy. Consequently, the word data were collapsed across this factor. The mean latency and accuracy for each condition is presented in Table 1 for words, and in Table 2 for pseudowords.

Words

As can be seen in Table 1, responses to frequent words were more accurate than responses to rare words. Due to the low variance for frequent words, further analyses for word accuracy were performed only on the rare words.

Presentations and word frequency. As expected there was an overall effect of repetition in latency, $F\ (1,30) = 92.74$, $p < .001$, and in accuracy, $F\ (1,30) = 29.09$, $p < .0001$, where responses to repeated words were both faster and more accurate than responses to experimentally novel words. There was also a main effect of word frequency, $F\ (1,30) = 102.10$, $p < .0001$, as well as the expected interaction for word frequency by repetition, $F\ (1,30) = 10.67$, $p < .003$. The effect of a prior presentation reduced response latency significantly for both frequent words (562 vs 525 msec for first and second presentation, $t\ (16) = 3.57$, $p < .05$; and rare words (643 vs 575 msec for first and second presentation, $t\ (16) = 3.57$, $p < .001$. Thus, a prior presentation benefited decisions about rare words more than decisions about frequent words. This result replicates similiar results by Scarborough, Cortese and Scarborough (1977) and Stanners, Jastrzembski and Westbrook (1975).

Instructions. The unbiased and pronunciation-biased group did not differ in their overall response latency or accuracy to words. However, there was a significant interaction for latency between groups and word frequency, $F(1,30) = 10.07$, $p < .003$, such that responses to rare words in the pronunciation-biased group were delayed by 49 milliseconds relative to the unbiased group (633 vs 584 msec, respectively). As can be seen in Table 1, a major portion of this delay is apparently due to judgments about abstract rare words; this was not significant, however, $F(1,30) = 1.61$, $p > .2$. The interaction of the secondary task with word frequency suggests that phonemic information is not typically used in making a lexical decision, and that the extra requirement to pronounce the word served as a secondary task similar to those discussed by Becker (1976). Although there was an interaction between groups and repetition (new and old words), $F(1,30) = 7.75$, $p < .01$, there was no overall difference in accuracy between the unbiased and the pronunciation-biased group (91% vs 92% accuracy, respectively). A post hoc analysis of the interaction showed that a prior presentation had a significant effect on accuracy in the unbiased group (83% vs 98% correct, for novel and repeated words; $p < .05$ by Sheffe's). The mean accuracy differed in the same direction in the pronunciation-biased group, (89% vs 94% correct) for experimentally novel words and repeated words. There were no other interactions for word decisions involving the secondary task.

Meaning. There was the expected effect of meaning, $F(1,30) = 24.16$, $p < .001$, as well as the expected interaction of meaning and repetition, $F(1,30) = 6.54$, $p < .015$, which revealed that the latency difference between concrete and abstract words was significant for experimentally novel words (584 vs 621 msec, for concrete and abstract words), but not for repeated words (544 vs 555 msec, for concrete and abstract words). In contrast to the results reported by James (1975), there was no significant interaction between word

TABLE 2

Reaction times (in msec) and Percent Accuracy for Pseudowords in Experiment 1.

CONTROL GROUP

	BISYLLABIC RT	BISYLLABIC ACC.	MONOSYLLABIC RT	MONOSYLLABIC ACC.
NEW	647	(91.5)	630	(94.1)
OLD	680	(83.4)	640	(90.9)

EXPERIMENTAL GROUP

	BISYLLABIC RT	BISYLLABIC ACC.	MONOSYLLABIC RT	MONOSYLLABIC ACC.
NEW	806	(89.4)	733	(95.0)
OLD	745	(85.4)	712	(90.6)

frequency and meaning, although there was a trend in this direction for novel words. The latency differences between concrete and abstract words in the present experiment were smaller than those reported by James, and the interaction between meaning and frequency may have failed to attain significance because of a floor effect. In general, the results support James' finding that meaning is a critical dimension in the Lexical Decision Task for novel or unfamiliar words, but not for familiar words.

Pseudowords

Instructions. In contrast to words there was a main effect of task instructions for pseudoword latency, F (1,30) = 7.68, $p < .01$, which revealed that response latency in the pronunciation-biased group (749 msec) was delayed relative to latency in the unbiased group (649 msec). The secondary task instructions to pronounce the stimuli delayed pseudoword responses by 100 msec. However, as there was no difference in the response accuracy for the two groups, performing the secondary task added little information relevant to a lexical decision. Hence, as for word decisions, the pseudoword decisions show that the secondary task affected decision time but not decision accuracy.

Repetition. There was a significant effect of a prior presentation, F (1,30) = 15.68; $p < .0005$, with repeated pseudowords being rejected less accurately than novel pseudowords (87% vs 92% accuracy, respectively). While there was no main effect of repetition on latency, there was an interaction involving repetition and task instructions in latency, F (1,30) = 12.30, $p < .002$, in which a prior presentation delayed responding in the unbiased group by 21 msec, but facilitated responding in the pronunciation-biased group by 40 msec. In light of the main effect of repetition upon errors, it is apparent that the increased latencies in the pronunciation-biased group reflect processing that was in addition to the processing required for a lexical decision. It seems reasonable to assume that the additional time reflected the time to generate the pronunciation. Moreover, in the pronunciation-biased group the increased familiarity from a prior presentation facilitated responses, whereas in the unbiased group increased familiarity delayed responding. A possible interpretation of this result is that the second pronunciation of a pseudoword was facilitated by the first pronunciation and this benefit overshadowed the negative benefits of familiarity in making a lexical decision about a repeated pseudoword. That is, information about the repeated pseudowords in the unbiased group interfered with the ability of the subjects to correctly reject the pseudowords for both latency and accuracy, while in the pronunciation-biased group information about the repeated pseudowords both interfered with the lexical decision and helped in the generation of a pronouncation for that pseudoword.

Syllable Length. In pseudowords, syllable length showed a significant main effect in response latency, $\underline{F}$ (1,30) = 56.07, $\underline{p} < .001$, and in response accuracy, $\underline{F}$ (1,30) = 43.62, $\underline{p} < .0001$, with bisyllabic pseudowords showing longer latencies (719 vs 678 msec) and lower accuracy (87% vs 92%) than monosyllabic pseudowords. That is, it was more difficult to reject as words bisyllabic pseudowords than monosyllabic pseudowords.

There were two interactions in the latency data which involved syllable length, one with task instructions, $\underline{F}$ (1,30) = 4.61, $\underline{p} < .04$, and a three way interaction with repetition and task instructions, $\underline{F}$ (1,30) = 8.20, $\underline{p} <$ .01. In general, the effect of syllable length was larger in the pronunciation-biased group (775 vs 723 msec for bisyllabic and monosyllabic pseudowords, respectively) than in the unbiased group (663 vs 634 msec for bisyllabic and monosyllabic pseudowords). However, this was modified by the interaction with presentation. Table 2 shows the mean response times for this three way interaction. A post hoc analysis showed that the effect of syllable length was significant for all comparisons (at $p < .01$, by Scheffe's) except that of novel pseudowords in the unbiased group. Of further interest, while a prior presentation increased the size of the syllabic effect in the unbiased group, it decreased the size of the syllabic effect in the pronunciation-biased group. Thus, once more the secondary task showed an effect of phonemic information opposite to that in the unbiased group. In the unbiased group, syllabic structure was significant only for repeated pseudowords, while in the pronunciation-biased group syllabic structure decreased in size for repeated pseudowords and was significant in all comparisons. A possible interpretation of this result in latency is that in the pronunciation-biased group bisyllabic pseudowords, being more difficult to pronounce, benefited from a prior pronunciation more than the monosyllabic pseudowords. For the unbiased group, on the other hand, syllabic structure became more important as the difficulty of rejection increased. This, perhaps, reflected a greater reliance on phonemic information as the difficulty of the decision increased. Yet, in light of the similar errors for the two instruction groups, it is not clear what benefits this strategy might provide. In conclusion, the secondary task of pronouncing the stimulus as well as making a lexical decision to it inverted the effect of syllable structure in pseudowords; this reversal makes sense when phonemic and lexical processing are considered as separate tasks.

The results of Exp. 1 provide information about the importance and the use of three different types of information in the Lexical Decision Task, phonemic, semantic, and visual information. The use of a secondary pronunciation task influenced the speed but not the accuracy of lexical decisions. The secondary task did not simply add a constant to the overall level of difficulty. The changes in response latency were contingent upon the type of stimulus, showing no change with frequent words, a small delay with rare words, and the largest delay for for pseudowords. This result, taken with the interaction of syllabic structure and the secondary task, suggests that phonemic information is not actively genertated in the typical Lexical Decision Task. More importantly, there was no change in accuracy associated with the increased latencies. This implies that phonemic information does not provide novel or distinctive information for a lexical decision.

The present results essentially replicate the findings of James (1975), and support the claim that semantic information is a potential source of information for subjects making a lexical decision. Moreover, the effect of meaning is most noticeable in decisions for experimentally novel words and is less important in decisions for repeated words. This supports a description of the Lexical Decision Task in which decisions are based upon memory as a whole, using information from a variety of sources if information is limited ~~at at~~ any one level. This issue will be discussed in detail in the following experiment.

The evidence for visual information ~~in the~~ Lexical Decision Task comes from two sources. First, the decreased reliance on meaning for repeated words seems to demand an increased reliance on information at a lower level. Visual information, given the small effects observed for phonemic information, seems to be the best candidate for this lower level of information. More important, perhaps, is the replication of the data reported by McKoon and Ratcliff (1979), Balota and Neely (1980) and Carroll and Kirsner (1982) who found that decisions for familiar pseudowords take longer and produce more errors than decisions for novel pseudowords. In addition, the present finding gives greater generality to the results reported McKoon and Ratcliff (1979). In their experiment, repeated pseudowords were learned as members of a paired associate list with a study time of 4 sec per pair, ten nonwords were learned per list, and only nonwords preceded by their paired associate were inhibited in the following Lexical Decision Task. In the present study forty pseudowords were seen only briefly as targets in a Lexical Decision Task, and were tested on the average 20 min later. Yet the repeated pseudowords led to an increase in errors and in the nonbiased group an increase in the time for correct rejection. The most likely information supporting these results is a memory for the prior occurrence of the pseudoword. Moreover, the most likely type of memory information supporting this increased familiarity is information about the visual encoding. Hence, it must be concluded that the information used in the Lexical Decision Task can be based upon a memory for a single instance of a visual target, and that not only meaningful words are retrieved in making lexical decisions. In summary, the Lexical Decision Task is not a test of vocabulary units abstracted over multiple experiences with words into a set of recognition units specific to language but rather, in part, it must test for any memory about previously experienced grapheme patterns.

Models of Lexical Decision

Effects of Meaning. The Verification Model in its present form cannot account for an effect of meaning in lexical decision. Presumably, it could explain the speeded responses for concrete words by assuming that the order of verification is not only biased in terms of word frequency but is also biased for meaning. Words with concrete referents would be entered for verification before words with abstract referents. How the type of referent would be coded for selection is unclear. Of course, the Verification Model could be restricted to memory look-up while an extra process was added to deal with the meaning of words. In either case, the Verification Model must be modified to account for the effects of meaning found here and by James (1975).

Morton's Logogen Model is quite explicit about the separation of meaning and the Visual Input Logogens. If a decision employs semantic information, then the Cognitive System must be involved in the decision. Thus, given the experimental evidence implicating semantic attributes in the Lexical Decision Task, lexical decisions must involve Morton's Cognitive System as well as the Visual Logogen System. Moreover, as the increased false alarms for repeated pseudowords require that response selection occur at the level of the graphemic representation (i.e., Logogen System) as well as the Semantic System, then lexical decisions must occur on the basis of the output in either of the two response systems, the Visual Logogen System and the Semantic System. Hence, in the Logogen System it must also be concluded that the Lexical Decision Task does not represent a pure measure of a single psychological process, or response set.

In lexical store models assume that subjects have a predetermined set of responses (i.e., a vocabulary of words). It is this restricted set of response units that supports a special lexical memory. The increased errors seen for repeated pseudowords undermines this assumption of a vocabulary based lexical memory, and argues instead for a visually based set of memories. Rejecting a pseudoword as a word is unlikely to make the pseudoword a meaningful vocabulary member and yet this simple visual task leads to a word bias in lexical decisions. By analogy, it may not be necessary to invoke memory constraints derived from vocabulary knowledge to explain word decisions.

One argument that might preserve the lexical store assumptions would be to assume that all incorrect word responses (i.e., errors to pseudowords) are due to confusions with real words. It could be that errors for repeated pseudowords were due to the same type of confusions that produce errors for novel pseudowords, namely a spelling confusion with a word. If this is so, then it must be that the visual information specified by a word detector is modified by the prior presentation of a similar pseudoword. In this case, a detector would not be a stable response unit as it is assumed, rather it would be a highly unstable unit whose input characteristics were continually being updated. By this argument, an increase in errors would be explained by the assumption that when subjects have incorrectly classified a pseudoword as a word in Phase I, they persevere with this spelling confusion in Phase II, and in addition experience novel word confusions among the repeated pseudowords in Phase II. That is, with exposure to similar alternatives word knowledge deteriorates. This perseveration hypothesis was tested by looking at the consistency of the errors produced in Phase I and Phase II. The analysis revealed that of the errors in Phase I, 53% were also errors in Phase II, while 47% were correctly identified as pseudowords in Phase II. These proportions do not differ by Sign Test from chance , $\underline{Z}$ = .7, N=167. In short, there is no evidence that the increased errors seen for repeated pseudowords are based on specific word confusions. Moreover, the decreased response times for repeated pseudowords in the pronunciation-biased group suggests that the pseudowords were encoded as discrete events, in order to provide an individualized pronunciation.

Effects of Repetition. The Relative Familiarity Judgment model does not restrict lexical decisions to a particular information level. Moreover, as it assumes that information is processed in time and is hierarchically constrained, then increasing the strength of lower level information should reduce the importance of higher order constraints. Hence, interactions between meaning and visual familiarity are predicted by the model.

The present pattern of results for repeated pseudowords, where accuracy and latency are inversely related, is consistent with, and indeed is predicted by, the description of lexical decision in terms of a Relative Familiarity Judgment. However, the observed increase in latency for repeated pseudowords is not consistent with the Verification Model, nor is the observed decrease in accuracy for repeated pseudowords consistent with the Logogen System. This issue will be discussed in more detail following Experiment 2.

Experiment 2

In Experiment 1 it was found that lexical decisions are sensitive to the semantic properties of words. It would be useful, however, to gain additional evidence of semantic processing in the Lexical Decision Task. Ideally, evidence in which semantic information is manipulated within an experiment. The Levels of Processing task (Craik & Lockhart, 1972) provides this experimental control of the information associated with a word.

The Levels of Processing manipulation entails the presentation of a word in one of several orienting tasks, such as, visual letter search, rhyme judgment and semantic judgment tasks. The typical result is that recall and recognition are contingent upon the type of orienting task in which a word was seen. Retention is usually best for semantic processing and poorest for visual search. Jacoby and Craik (1979) suggested that this difference in retention is due primarily to the different types of information encoded during the orienting tasks, rather than in a simple increase in the overall strength of the item in memory. Jacoby and Dallas (1981) extended this analysis by showing that the facilitation usually seen with a prior presentation in word detection is independent of the type of encoding task, while the typical levels of processing effect is present in tests of recognition memory. Jacoby and Dallus (1981) suggest that as word identification is primarily based upon visual information, and as all orienting tasks require visual processing, then similar benefits in visual processing are to be expected for any orienting task using a visual presentation. In other words, the processing during a test trial will select from a memory episode only the information that is relevant to the task. This provides a technique for measuring the type of information required by the Lexical Decision Task. If, like the Word Identification Task, the Lexical Decision Task is primarily based upon visual information, then the type of orienting task used during study trials should not lead to different benefits of repetition. However, if the Lexical Decision Task uses semantic as well as visual information then the different orienting tasks should have different benefits in the Lexical Decision Task, with semantic judgments typically providing greater benefits than physical judgments.

An additional consideration in Exp. 2, came from an experiment reported by Balota and Neely (1981) using the Levels of Processing task. In their experiment they found a larger benefit in a Lexical Decision Task with words previously seen for semantic judgment, than with words seen for rhyme judgment. A problem exists with their data, however, in that while response latencies show a benefit for a prior semantic encoding, this advantage was offset by a decrease in the accuracy for these items. Moreover, the size of the facilitation in latency, approximately 20 msec, did not change with a between subjects manipulation in the ease of making a lexical decision. In their experiment lexical decisions were made at two levels of difficulty, a difficult decision which used pronounceable nonwords as distractors, and an easy decision which used unpronounceable nonwords as distractors. As would be expected, decisions using the unpronounceable nonwords were both faster and more accurate than decisions using pronounceable pseudowords, yet the Phase I manipulation of semantic encoding was slightly larger for decisions using unpronounceable nonwords.

The results reported by Balota and Neely (1981) do not support the present argument for an increased dependence on semantic constraints as decision difficulty increased. Given the results of Balota and Neely (1981), it seemed necessary to demonstrate that with a change in task demands the identical manipulation of prior encoding context both would, and would not, result in differences of transfer in a following Lexical Decision Task. This demonstration of flexible processing in the Lexical Decision Task, would serve two purposes. First, it would emphasize the role of different types of information in the Lexical Decision Task and second, it would demonstrate the influence of task demands upon the types of information used in a particular Lexical Decision Task.

A between subjects manipulation of task instructions was chosen to modify task demands. All subjects were to be given the same Levels of Processing task, which involved presentation of words and pseudowords in one of two orienting tasks, a visual letter search or semantic judgment task. In a following Lexical Decision Task subjects in one group of subjects were asked to respond quickly, while subjects in the other group were asked to respond accurately. This constituted the instruction manipulation. In the Relative Familiarity Judgment model the effect of the speed and accuracy instructions should be to change the value of the information criterion used for decisions.

That is, the instructions should change the decision boundary (A) in the random walk process. Subjects who received the speed instructions should adopt a small or liberal value. The use of a low decision boundary will lead to faster responses, at the cost of an increase in errors. However, errors will not increase uniformly. Errors will increase most for those items with the lowest absolute slope, i.e., U_2 and V_1 in Figure 3. Hence, it is predicted that decisions for rare words and repeated pseudowords will be hurt most by a low decision barrier. This prediction differs from that given by the lexical store models. Both the Logogen System and the Verification Model predict that with speeded decisions errors for rare words will increase relative to frequent words and, as both theories must consider repeated pseudowords to be instances of very rare words, the lexical store models predict that errors to repeated pseudowords will decrease relative to rare words.

Another aspect of adopting a low decision boundary in the random walk process is that if different types of information have different mean arrival times then the types of information sampled for a decision will be a function of the boundary. Decisions should be less sensitive to the semantic properties of words with a low decision boundary. It is suggested, then, that performance in the speed instruction condition should be mediated mainly by visual familiarity. The opposite predictions are made for subjects who receive the accuracy instructions. That is, they should adopt a high or conservative decision boundary and decisions should be delayed and more accurate. Also, stimuli are more likely to be elaborated in terms of non-visual properties. This should result in a largec influence of semantic information, and in very few errors to repeated pseudowords.

The degree of semantic processing was further assessed in terms of responses to words which differed with regard to their semantic referent, concrete or abstract. Similar to Exp. 1, decisions about rare words can be expected to show a larger benefit of meaning than decisions about frequent words. Frequent words are likely to be sufficiently familiar to be responded to on the basis of visual information alone.

The manipulation of encoding context, using the Levels of Processing task, may be more important for abstract words than for concrete words. This can be expected for two reasons. First, by analogy to word frequency and repetition, it can be expected that difficult decisions will benefit from additional information more than easy decisions. Second, information about concrete and abstract words may be retrieved in a different fashion. Referents of concrete words are easily imaged and retrieval may entail elabboration of this information. Abstract words, on the other hand, may depend more on the reinstatement of a prior episode.

To summarize, it was predicted that the manipulation of speed and accuracy instructions would lead to changes in the type of information used in the Lexical Decision Task. The use of speed instructions should force decisions to rely on visual information and as a result 1) decisions should reflect only a small influence of semantic information; and 2) decisions to repeated pseudowords relative to novel pseudowords should exhibit a word bias. The use of accuracy instructions should induce some form of meaning check in addition to the use of visual information, and hence 1) decisions for difficult items should reflect semantic information; and 2) decisions about repeated pseudowords should be no less accurate than decisions about novel pseudowords, although because of their greater visual familiarity repeated pseudowords may take longer to reject than novel pseudowords.

The effect of the Levels of Processing manipulation on recognition memory was not tested in the present experiment for two reasons. First, in a pilot experiment reported in Appendix A, a test of recognition memory showed the expected effect of the type of encoding, while there was no significant effect in a Lexical Decision Task. Secondly, Balota and Neely (1981) demonstrated a large levels of processing effect on recognition, while finding only a small effect of type of encoding in lexical decisions. As the present manipulation of encoding tasks was similar to these previous manipulations of encoding which produced a large effect of encoding task on recognition memory, it was considered unnecessary to test the effect of the encoding task on recognition memory in the present experiment.

Method

<u>Subjects</u>. Sixty four McMaster undergraduates were tested as part of a course requirement. Each subject was assigned randomly to one of two groups, a speed instruction or an accuracy instruction group.

<u>Materials</u>. A total of 128 stimulus words were selected from Thorndike-Lorge, of which 64 were high frequency A words and 64 were low frequency words (1-13 occurrences per million, with an average frequency of 4.25). These words were selected with the constraint that they be 5 or 6 letters in length, and maximized a division between concrete and abstract properties. Concrete words were operationally defined to be words with a referent which can be directly perceived in one of the sensory systems (i.e, KNIFE, BRANCH, ANVIL or SPLEEN); while an abstract word was anything that did not fit this classification (i.e., GRAND, CHOICE, BLAND or TRANCE. Of these examples, the first two words are frequent words while the latter two are rare words.). Thus, there were 32 words in each of the four word categories - 2

frequency times 2 referent categories. In addition, 128 pseudowords were constructed from medium frequency, 5 and 6 letter monosyllabic and bisyllabic words by replacing one or two lettters (i.e., SKILL -> SKALL : FOURTH -> POURTH), with the constraint that no pseudoword should be directly confusable in spelling with a word (i.e., SKILL -> SKELL), nor was an obvious homophone for an English word. The words and pseudowords were partitioned into 8 sets, of 16 words and 16 pseudowords, for purposes of counterbalancing, with each set of words containing 4 words from each of the four word classes. These items were rotated over subjects so that each word appeared in each experimental condition equally often.

<u>Design and Procedure</u>. Each subject was randomly assigned to one of two groups, the speed instruction and accuracy instruction groups, with the provision that each group contained 32 members. An experimental session consisted of two phases, a Phase I, Levels of Processing manipulation; and a Phase II, Lexical Decision Task. The subjects in each group were given the same Phase I manipulation, but differed in the task instructions they received for the Phase II Lexical Decision Task. The Phase I manipulation required subjects to perform two tasks, a visual letter search (e.g., Contains the letter <u>B</u>) and a semantic judgment task (e.g., Is a RODENT?) by answering with a <u>yes</u> or <u>no</u> as quickly as possible. Each subject was given 32 words in each orienting task for a total of 64 words, 48 pseudowords in the visual search task, and 16 pseudowords in the semantic judgment task (there were no semantic <u>yes</u> questions for the pseudowords) for a total of 64 pseudowords. Each subject received 8 words for each of the four word classes, in each orienting task. An equal number of <u>yes</u> and <u>no</u> responses were required for each task.

Phase II was a Lexical Decision Task and contained 128 word stimuli and 128 pseudoword stimuli, of which 64 words and 64 pseudowords had been seen in Phase I. At the beginning of the Phase II task subjects were given the appropriate task instructions for their group, followed by 20 practice trials. Speed instruction subjects were asked to respond as quickly as possible in making a lexical decision, and to basically respond on the basis of their first impression. Accuracy instruction subjects were asked to respond as accurately but as quickly as possible in making a lexical decision. In addition, subjects receiving accuracy instructions were given error feedback in the form of a soft buzz whenever they made an incorrect response.

All aspects of the experiment were controlled and sequenced by a DIGITAL PDP-8A computer, which collected and recorded the responses to the nearest millisecond. The presentation procedure was identical to that used in Exp. 1.

Results and Discussion

A median latency and mean accuracy score was computed for each condition and each stimulus type for each subject. The data were analyzed separately by an analyses of variance with latency and accuracy as dependent measures.

Table 3 displays the main effects of the between groups instruction manipulation. As can be seen, there was a large and significant effect of instructions, with accuracy instructions producing longer latencies (621 vs 532 msec) and fewer errors (6.2% vs 10.2%) than speed instructions. This response pattern is easily interpreted in terms of the predicted speed - accuracy trade off. As this pattern of results is present across all subsequent comparisons, it will not be mentioned further. The following analysies were computed

TABLE 3

Mean Group Reaction Times (in msec)
and Percent Accuracy for
the Speed and Accuracy Instruction Conditions
in Experiment 2.

	WORDS		PSEUDOWORDS	
	RT	ACC.	RT	ACC.
ACCURACY INSTRUCTIONS	621	(93.8)	742	(92.3)
SPEED INSTRUCTIONS	533	(89.8)	628	(85.4)

separately for words and pseudowords.

The effect of a prior presentation was analyzed by two comparisons. First, lexical decision responses to old items were compared across the two orienting tasks, and secondly lexical decision responses to new items were compared to all old items collapsed across the Phase I orienting task.

Responses to Repeated Stimuli

Repeated Words. Table 4 presents the mean-median latency and accuracy for the repeated words, contingent upon the Phase I, Levels of Processing manipulation, visual letter search or semantic judgment. In the following results only the data for rare words will be reported as the responses to all frequent words were near their performance ceiling and contained few difference between the experimental conditions. In all cases though, the direction of the differences for frequent words paralleled those for rare words.

The data were analyzed by a (2 X 2 X 2 X 2) analyses of variance, comparing instruction groups (speed or accuracy), type of orienting task (letter search or semantic judgment), type of response (yes or no), and type of word referent (concrete or abstract).

The instruction groups differed reliably in overall latency, F $(1,62) = 15.6$, $p < .001$, and accuracy, F $(1,62) = 5.6$, $p < .02$, with responses in the accuracy instruction condition being both slower (642 msec) and more accurate (93.4%) than responses in the speed instruction condition (554 msec and 88.1%, respectively). More interestingly the Phase I, orienting task produced a significant effect for both latency, F $(1,62) = 5.45$, $p < .025$ and accuracy, F $(1,62) = 13.63$, $p < .001$, with words previously seen in the semantic judgment task responded to faster (605 vs 591 msec) and more accurately (93.1% vs 88.4%) than words previously seen in the letter search task. Additionally, there also

54a

TABLE 4

Mean-Median Reaction Times (in msec)

and Percent Accuracy for

Repeated Words, in Experiment 2.

ACCURACY INSTRUCTIONS

Phase I TASK	RESP.	HIGH FREQUENCY WORDS: CONCRETE RT	ACC.	ABSTRACT RT	ACC.	LOW FREQUENCY WORDS: CONCRETE RT	ACC.	ABSTRACT RT	ACC.
LETTER	N:	554	(99.2)	563	(96.1)	625	(95.4)	705	(84.4)
LETTER	Y:	558	(100.)	579	(98.4)	600	(97.7)	681	(90.6)
SEMANTIC	N:	544	(99.2)	569	(96.1)	635	(98.4)	655	(89.8)
SEMANTIC	Y:	542	(100.)	578	(99.2)	605	(95.3)	631	(95.3)

SPEED INSTRUCTIONS

Phase I TASK	RESP.	HIGH FREQUENCY WORDS: CONCRETE RT	ACC.	ABSTRACT RT	ACC.	LOW FREQUENCY WORDS: CONCRETE RT	ACC.	ABSTRACT RT	ACC.
LETTER	N:	482	(100.)	487	(94.5)	551	(87.5)	582	(77.3)
LETTER	Y:	475	(99.2)	482	(96.9)	539	(90.6)	557	(83.6)
SEMANTIC	N:	465	(99.2)	487	(96.9)	556	(91.4)	566	(87.5)
SEMANTIC	Y:	472	(100.)	485	(97.7)	517	(96.1)	566	(90.6)

Where N = NO, and Y = YES responses in Phase I.

a main effect of the type of response for both latency, F (1,62) = 7.65, p< .01, and accuracy, F (1,62) = 10.40, p< .002, where words previously responded to with a yes response were both faster (587 vs 609 msec) and more accurate (92.5% vs 89%) than words previously responded to with a no response. In summary, the Phase I manipulations had a significant effect on a subsequent Lexical Decision Task in both decison latency and accuracy. The current effects of the type of Phase I manipulation and Phase I response contradicts the generalized knowledge assumptions of lexical store models.

Meaning. As seen in Exp. 1, meaning was highly significant for both latency, F (1,62) = 28.79, p< .001, and accuracy, F (1,62) = 25.78, p< .001. Of especial interest there were two interactions involving meaning in the response latency, a marginal two way interaction with orienting task, F (1,62) = 4.68, p< .04, and a three way interaction with instruction condition by orienting task, F (1,62) = 6.52, p< .015. As both factors in the two way interaction are included in the three way interaction, only the three way interaction will be discussed.

The reaction times and the percent accuracy for each cell of the three way interaction are presented in Table 5. The interpretation of this interaction is clear. In the table, it can be seen that the source of the interaction resides in the responses to abstract words. Specifically, the Phase I orienting task changed performance with abstract words in the accuracy instruction condition more than in the speed instruction condition. Consequently, judgments about abstract words were helped by the prior semantic context more in the accuracy condition, than in the speed condition. Out of four possible comparisons involving the concrete and abstract words, there was only one comparison where the effect of the type of orienting task was significant at $p < .05$ (by Scheffe's Test), and this was for the abstract words in the accuracy instruction condition where the latency for judgments about abstract words previously seen in the Phase I, semantic judgment task are 49 milliseconds faster than the same words seen in the Phase I, letter search task (693 vs 644 msec, respectively). The analyses of the latencies for abstract and concrete words, in Table 5, revealed that all four comparisons were significant at $p < .01$, while the latency difference in the accuracy by letter search comparison (81 msec) was significantly different from all other cells with $p < .001$. In conclusion, there was a small but significant effect of meaning in all comparisons, but the effect of meaning interacted with both the prior encoding context and the current task demands.

Although the reaction time data were easily interpreted and there were no significant interactions in the accuracy of responding, the variability in accuracy across the eight cells in Table 5 suggested that there might have been some trade-off present between the accuracy and latency of responding. To examine this possibility, the marginal probability was computed for each of the conditions in Table 5; these marginals were then used to compute an expected score for each of the eight cells. When the expected and the obtained values were compared, none of the obtained scores deviated from the expected score by more than 1.5%, and the Chi Square statistic computed for the eight comparisons was .29, which is not significantly different from chance. Further, the pattern of the computed difference scores was in the opposite direction to that required to explain the latency difference found in the three-way interaction in terms of a speed - accuracy trade off.

In summary, the experimental manipulation of semantic information revealed a positive benefit of semantic information in the lexical decision

TABLE 5

Mean-Median Reaction Times (in msec)

and Percent Accuracy for

the Instructions * Task * Meaning Interaction

in Experiment 2.

MEANING:		CONCRETE WORDS			ABSTRACT WORDS		
GROUP		PHASE I TASK			PHASE I TASK		
		LS	SJ	L-S	LS	SJ	L-S
ACCURACY INSTRUCTIONS	RT	612	620	-8	693	644	49
	ACC	96%	97%		87%	93%	
SPEED INSTRUCTIONS	RT	545	537	8	570	566	4
	ACC	89%	94%		80%	89%	

Where: LS = LETTER SEARCH TASK, IN PHASE I.

SJ = SEMANTIC JUDGEMENT TASK, IN PHASE I.

times. However, this benefit was seen only for unfamiliar, abstract words and only when subjects were trying to be as accurate as possible. These results are in agreement with the experimental predictions. Access to meaning for concrete words was assumed to be generally independent of context, and hence concrete words were not expected to show specific benefits based upon the encoding context. Access to meaning with abstract words, on the other hand, was assumed to be relative to some context of interpretation, since abstract words are interpreted as specifying a relationship. Hence, judgments about abstract words were expected to be helped by a prior presentation of a specific context. The results support these predictions. Provision of context, the semantic judgment task in Phase I, speeded lexical decisions for abstract words while judgments about concrete words appeared to be largely independent of the type of prior presentation.

Repeated Pseudowords. Table 6 presents the mean-median latency and mean accuracy scores for these comparisons. There was no effect of the Phase I orienting task in either the latency or accuracy of pseudoword decisions. It would appear that the context in which a pseudoword has been encountered has little effect on processing. Merely performing a letter search in a pseudoword string is sufficient processing to render the pseudoword a familiar stimulus.

There was a main effect of instructions in latency, $\underline{F}$ (1,62) = 16.08, $\underline{p} < .001$, and in accuracy, $\underline{F}$ (1,62) = 7.48, $\underline{p} < .01$, with pseudowords in the speed instruction condition being faster and less accurate (643 msec & 82.6%) than pseudowords in the accuracy condition (759 msec & 90.5%) thus supporting the generally poorer performance in the speed condition. There was a main effect of pseudoword length in latency, $\underline{F}$ (1,62) = 51.88, $\underline{p} < .001$, and in

TABLE 6

Mean-Median Reaction Times (in msec) and Percent Accuracy for Repeated Pseudowords, in Experiment 2.

ACCURACY INSTRUCTIONS

Phase I TASK	RESP.	MONOSYLLABIC PSEUDOWORDS 5-LETTER RT	ACC.	6-LETTER RT	ACC.	BISYLLABIC PSEUDOWORDS 5-LETTER RT	ACC.	6-LETTER RT	ACC.
LETTER	N:	724	(96.1)	795	(93.0)	732	(93.4)	793	(88.3)
LETTER	Y:	718	(94.5)	759	(90.6)	765	(93.0)	782	(80.5)
SEMANTIC	N:	708	(95.3)	791	(88.3)	747	(91.4)	790	(81.3)

SPEED INSTRUCTIONS

Phase I TASK	RESP.	MONOSYLLABIC PSEUDOWORDS 5-LETTER RT	ACC.	6-LETTER RT	ACC.	BISYLLABIC PSEUDOWORDS 5-LETTER RT	ACC.	6-LETTER RT	ACC.
LETTER	N:	625	(88.3)	664	(82.8)	626	(89.8)	672	(72.7)
LETTER	Y:	624	(89.1)	669	(85.3)	633	(86.7)	641	(79.7)
SEMANTIC	N:	628	(87.5)	688	(77.3)	608	(86.7)	633	(72.6)

Where N = NO, and Y = YES responses in Phase I.

accuracy, $\underline{F}$ (1,62) = 90.01, $\underline{p}$< .001, with 5-letter pseudowords being responded to faster and more accurately (678 msec & 91.1%) than 6-letter pseudowords (723 & 81.7%). Also, there was a main effect of syllable length in accuracy, $\underline{F}$ (1,62) = 13.26, $\underline{p}$< .001, with monosyllable pseudowords being responded to more accurately than bisyllabic pseudowords (88.1% vs 84.7%, respectively). In addition, there was a two way latency interaction which involved syllable length and accuracy instructions, $\underline{F}$ (1,62) = 8.76, $\underline{p}$< .005. This interaction showed that the effect of syllable length was reversed for the two instruction conditions. Whereas, monosyllabic pseudowords were responded to faster than bisyllabic pseudowords (749 vs 768 msec, respectively) in the accuracy condition, in the speed condition bisyllabic pseudowords were responded to faster than monosyllabic pseudowords (635 vs 650 msec, respectively). In light of the similar effects of syllabic structure on response accuracy, it is not clear what to make of this interaction with instructions. However, a possible interpretation of this result is that subjects receiving accuracy instructions tended to indulge in verbal coding more than subjects receiving speed instructions.

<u>Comparison of Experimentally New and Repeated Stimuli</u>

The mean-median latency and accuracy for the words are listed in Table 7 for the eight conditions, while Table 8 lists the mean-median latency and accuracy for the eight pseudoword conditions.

The word responses were analyzed by a (2 X 2 X 2 X 2) analyses of variance (two groups, two presentations, two levels of frequency and two levels of meaning). The pseudoword responses were analyzed by a (2 X 2 X 2 X 2) analyses of variance (two groups, two presentations, two levels of pseudoword length, and two levels of syllable length).

TABLE 7

Mean-Median Reaction Times (in msec)
and Percent Accuracy for
OLD vs NEW Words, in Experiment 2.

ACCURACY INSTRUCTIONS

	HIGH FREQUENCY WORDS				LOW FREQUENCY WORDS			
MEANING:	CONCRETE		ABSTRACT		CONCRETE		ABSTRACT	
	RT	ACC.	RT	ACC.	RT	ACC.	RT	ACC.
NEW	583	(98.6)	606	(95.3)	675	(90.6)	745	(81.8)
OLD	546	(99.6)	565	(97.5)	604	(96.7)	643	(90.0)

SPEED INSTRUCTIONS

	HIGH FREQUENCY WORDS				LOW FREQUENCY WORDS			
MEANING:	CONCRETE		ABSTRACT		CONCRETE		ABSTRACT	
	RT	ACC.	RT	ACC.	RT	ACC.	RT	ACC.
NEW	490	(98.0)	513	(95.7)	582	(84.0)	638	(68.6)
OLD	467	(99.6)	481	(96.5)	528	(91.4)	559	(84.8)

TABLE 8

Mean-Median Reaction Times (in msec)

and Percent Accuracy for

OLD vs NEW Pseudowords, in Experiment 2.

ACCURACY INSTRUCTIONS

	MONOSYLLABIC PSEUDOWORDS				BISYLLABIC PSEUDOWORDS			
	5-LETTER		6-LETTER		5-LETTER		6-LETTER	
	RT	ACC.	RT	ACC.	RT	ACC.	RT	ACC.
NEW	703	(98.4)	748	(92.6)	706	(94.5)	777	(89.8)
OLD	710	(95.1)	773	(90.6)	735	(92.8)	786	(84.6)

SPEED INSTRUCTIONS

	MONOSYLLABIC PSEUDOWORDS				BISYLLABIC PSEUDOWORDS			
	5-LETTER		6-LETTER		5-LETTER		6-LETTER	
	RT	ACC.	RT	ACC.	RT	ACC.	RT	ACC.
NEW	605	(94.3)	627	(85.9)	588	(93.0)	661	(82.6)
OLD	623	(88.3)	664	(78.3)	614	(87.9)	647	(73.2)

<u>General Effects of Speed and Accuracy Instructions</u>. As in Exp. 1 the result of a prior presentation had an opposite effect for words and pseudowords. Repeated words were both significantly faster than experimentally novel words, $\underline{F}$ (1,62) = 109.81; $\underline{p}$< .001, and more accurate than new words, $\underline{F}$ (1,62) = 73.04; $\underline{p}$< .001. The facilitation in decision times of a word repetition was the same for both instuction groups.

The opposite pattern was found for the responses to pseudowords. Latencies for repeated pseudowords were signigicantly slower than latencies for novel pseudowords, $\underline{F}$ (1,62)= 14.08; $\underline{p}$< .001, while accuracy was decreased by a prior presentation, $\underline{F}$ (1,62) = 44.40; $\underline{p}$< .001. In addition, the predicted interaction between pseudoword repetition and instructions was significant for response accuracy, $\underline{F}$ (1,62) = 6.80; $\underline{p}$< .01. The decrease ~~in a~~ccuracy for repeated pseudowords was larger in the speed instruction condition than in the accuracy instruction condition. The mean percent corrrect for these conditions are presented in Table 3 and while the general effect of a repetition is the same for the two groups, the speed instruction group shows a greater decrease in accuracy due to a prior presentation. The increase in errors for a repeated pseudoword was significant in only the speed condition. It would appear that subjects in the accuracy condition are less likely to respond on the basis of increased visual familiarity. In either case it can be concluded that the relevant background for making a lexical decision is at least in part task dependent. This result makes sense only within a relative judgment type of model, and contradicts predictions made by the lexical store models. This will be discussed in more detail in the following general discussion.

<u>Words</u>. There was one interaction involving the instruction groups in the repeated word data. This was for response accuracy, and involved word

frequency, $\underline{F}$ (1,62) = 13.21; $\underline{p} < .001$. While, the response accuracy for frequent words (97.8% vs 97.5%, for accuracy and speed instructions) was near a ceiling level of performance for both groups, the accuracy for rare words was significantly different for accuracy and speed instructions (89.8% vs 82.2%, respectively; $\underline{p} < .05$ by Scheffe's). Thus while the experimental instructions produced a main effect in decision latency for frequent and rare words (decision latency for both word types was decreased by a similar amount in the speed condition from that seen in the accuracy condition), these instructions produced dissimilar results in decision accuracy. This is a surprising result, as typically decision latency and decision accuracy are expected to vary together. Indeed, for reasons that will be discussed later, this result offers an important clue to the nature of lexical decisions.

<u>Other Results for Word Responses</u>. As was expected the effect of word frequency yielded highly significant results both in response latency, $\underline{F}$ (1,62) = 345.8, $\underline{p} < .001$, and accuracy, $\underline{F}$ (1,62) = 133.01, $\underline{p} < .001$. Word frequency interacted with the number of presentations in latency, $\underline{F}$ (1,62) = 32.40, $\underline{p} < .001$, and in accuracy, $\underline{F}$ (1,62) = 49.69, $\underline{p} < .001$, such that rare words benefited more from a prior presentation than did frequent words.

There was also a main effect of word meaning in both latency, $\underline{F}$ (1,62) = 106.2, $\underline{p} < .001$, and accuracy, $\underline{F}$ (1,62) = 83.31, $\underline{p} < .001$. Further, there was an interaction in latency between word meaning and repetition, $\underline{F}$ (1,62) = 32.40, $\underline{p} < .001$, with the effect of meaning being larger for new words (43 msec) than repeated words (24 msec); and an interaction in latency between word meaning and word frequency, $\underline{F}$ (1,62) = 23.05, $\underline{p} < .001$, with the effect of meaning larger for rare (49 msec) than for frequent (19 msec) words. In addition, there were parallel interactions in accuracy between word meaning and repetition, F (1,62) = 5.99, $p < .02$, and word meaning and word frequency, F (1,62) = 47.28, $p < .001$. There was also a three way interaction between word meaning, repetition and word frequency, F (1,62) = 47.28, $p < .015$, where there were no differences in accuracy for frequent words between concrete and abstract words, nor between repeated and new the more extensive processing included verbal processing. A similar explanation may explain the syllabic differences in Exp. 2 if it is assumed that subjects in the accuracy condition needed to process the pseudowords more extensively than subjects in the speed condition. This assumption, however, is not supported by the accuracy data, where for both instruction conditons there were more errors for bisyllabic than for monosyllabic pseudowords.

In conclusion, Exp. 2 like Exp. 1 yielded significant effects of pseudoword structure in the Lexical Decision Task, and as in Exp. 1 the latency measure interacted with other conditions, while there was a consistent main effect of pseudoword structure in response accuracy for both experiments.

Latency Distributions. Following the suggestion by Ratcliff (1978), that the shape of the response distribution can also provide relevant response information, the latency distributions for frequent and rare words, and for new and old words, were generated for both instruction conditions. The minimum and maximum correct response time was found for each subject in the four word conditions in each instruction condition. The relative frequency of a correct response was sorted into ten equal periods of time which covered the range between the mean minimum and maximum latency in each of the four conditions. These data will be discussed more fully in a following section.

General Discussion

The results of Exp. 2 are in good agreement with the predicted results. Lexical decisions about words were helped by a prior exposure to semantic information. This was true, however, only for certain words and only when subjects were motivated to be as accurate as possible. For frequent words and concrete words the manipulation of semantic information had no obvious effect. Decisions for frequent words are probably based solely on the familiarity of the visual grapheme, and typically show little effect of semantic information (James, 1975; Exp. 1). The failure to find an influence of prior semantic context with frequent words agrees with these earlier studies. While it was expected that meaning for concrete words might depend less on retieval of a particular prior encounter than meaning for abstract words, finding no benefit of a prior semantic context for rare, concrete words was somewhat surprising. It is possible that given a greater emphasis on accuracy a benefit of a prior context would be found.

The evidence for benefits of a prior semantic context in lexical decisions was seen in decisions for rare, abstract words. Decisions about abstract words were 50 msec faster when there was a prior semantic context, if subjects were given accuracy instructions. There was no effect of prior context when subjects were making speeded decisions. On the basis of this interaction it can be concluded that the use of meaning in making a lexical decision is an optional strategy or, at the very least, that the reliance on meaning in the Lexical Decision Task is modified by task demands. Nonetheless, there was a significant effect of concrete and abstract words in the speed condition. At least in part, then, subjects relied on semantic information when required to make a speeded lexical decision. Hence, whereas it was suggested that processes would drop out in speeded decisions, the data do not support such a strong position. Even though decisions were on average twenty percent faster, in the speed condition relative to the accuracy condition, the same qualitative variables seemed to be mediating lexical decisions in both groups. The difference seemed to be a change in the emphasis of information and not a change in type of information used.

Subjects receiving speed instructions made more errors for repeated pseudowords. This implies the presence of visually based information. Yet there was also a small but significant effect of word meaning for speed instructions. Hence, as in Exp. 1, it can be concluded that at least two types of information are needed to explain lexical decisions, lexical decisions draw upon multiple types of information.

An unexpected result in Exp. 2 was the overall increase in accuracy contingent upon orienting task. It is not clear whether this was an effect of the type of information encoded about a word or a failure to encode some study items as words. Repeated words when seen in the semantic judgment task were responded to more accurately than repeated words seen in the letter search task. This increase in accuracy occurred in all test conditions, and tended to be larger for abstract words than concrete words. The interaction in response accuracy between orienting task and meaning was not significant ($p < .065$), and in fact appeared to be largest in the speed instruction group. While this result in the absence of similar reaction time differences is difficult to interpret, one possible explanation comes to mind. Briefly, it is possible that some of the words presented during the letter search task were not interpreted as that word (i.e., cargo might have been structured in terms of its constituents car-go, etc.). There is some evidence that this can happen, i.e., Aderman and Smith (1971) and Carr, Davidson, and Hawkins (1978). Moreover, it is likely that this failure of incidental processing would occur more often for rare words than for frequent words.

The following sections discuss in detail the implications of Exp. 2 for the theoretical issues discussed in the introduction. It will be argued that neither the Verification Model nor the Logogen System can satisfactorally account for these results, while the Relative Familiarity Judgment model predicts the overall pattern of the results. In a following section the arguments used to explain the increased errors to repeated pseudowords will be shown to be similar in kind to the processes that typically facilitate the recognition of repeated words; and it will be argued that the repetition effect typically observed with words in the Lexical Decision Task is a general structural change, and is not specific to word knowledge, per se.

Pseudoword errors

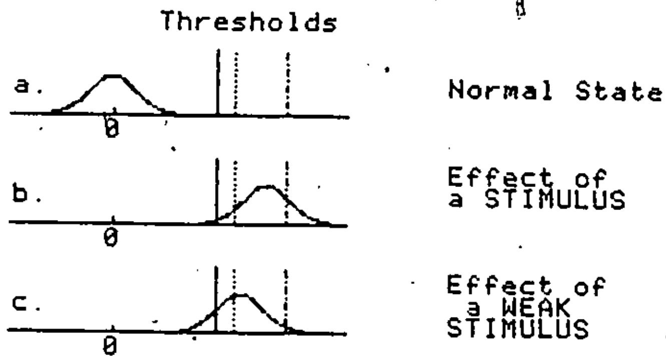
There were two main findings for pseudoword accuracy. First, in Exp.s 1 and 2 repeated pseudowords were shown to be more difficult to reject, both in terms of accuracy and latency than novel pseudowords. Secondly, in Exp. 2 errors for repeated pseudowords increased significantly when subjects were asked to respond quickly. Both of these results contradict the formal properties of the Logogen System and the Verification Model and, as will be shown, the results are not easily predicted by these theories even with major changes in the theoretical assumptions.

The pseudoword data replicate the results of Exp. 1. Errors for repeated pseudowords were more frequent than errors for novel pseudowords. Moreover, as predicted, this increase in errors was significant in only the speed instruction condition. As in the unbiased condition in Exp. 1, the decisions for repeated pseudowords were delayed relative to decisions for novel pseudowords. This was true for both instruction conditions, although there was no sign of the predicted increase in decision times for repeated pseudowords in the accuracy instruction condition. In short, while subjects in the accuracy instrucation condition were more accurate in rejecting repeated pseudowords, they did not achieve this at the expense of longer decision times. This pattern of results, where decreases in errors are not accompanied by increases in decision latencies, will be discussed in detail in the final discussion. For the present purposes, only the relative increase in errors for repeated pseudowords is of critical interest.

The pseudoword errors are important in determining how familiarity affects decisions about pseudowords. Earlier, it was mentioned that a prior presentation of a pseudoword might in effect turn the pseudoword into a very rare word. The results of Exp. 2 rule out this assumption, for the following reasons. As can be seen in Table 7, the effect of speed instructions compared to accuracy instructions was to decrease the accuracy for rare words more than for frequent words. This is consistent with simple threshold models, such as the Logogen System and the Verification Model which uses thresholds for a preprocessing stage. These theories assume an array of passive word detectors which vary in their thresholds. For these detectors the only difference between frequent and rare words is that detectors for frequent words require less information to be triggered than do detectors for rare words. This is shown schematically in Figure 7. As the information from the stimulus has a positive, non-zero normal distribution, then on average the probability that sufficient stimulus evidence is present for a frequent word response will be greater than the corresponding probability that sufficient evidence is present for a rare word response. In forcing a subject to respond quickly, as in the speed instructions in Exp. 2, we in effect decrease the average amount of stimulus

Fig. 7

The effects of certain situations, upon THE STATE OF A LOGOGEN.



Key to Thresholds:

——— = Frequent Words

········ = Rare Words

- - - - = Very Rare Words, or Repeated Pseudowords.

Where: the horizontal axes represents The level of excitation in the logogen. The curves correspond to probability distributions of the excitation. The vertical lines represent the thresholds of logogens.

When the level of excitation exceeds the threshold the corresponding word is available as a response.

evidence that is available for any response. As can be seen in Figure 7, this reduction in stimulus evidence, given the previous assumptions, will result in a greater loss of accuracy for rare words than for frequent words. The predictions are slightly different for very rare words (i.e., repeated pseudowords), as the criterion is located at the opposite end of the distribution and detection accuracy should not drop as quickly for low frequency words. Detection accuracy, however, should decrease with decreasing stimulus information. This is not what was observed with repeated pseudowords. Rather, errors to repeated pseudowords, which would correspond to correct detections of very rare words, actually increased with decreasing stimulus evidence. As a result, the increased errors observed for repeated pseudowords cannot be attributed to very rare word detectors formed by the prior presentation of the pseudoword within a word detection system. Moreover, it can also be concluded that accuracy in the Lexical Decision Task cannot be explained in terms of word detectors alone. That is, there must also be processes which are specifically involved in the rejection of distractors, such as that employed in the random walk model.

Latency and Accuracy Responses in the Verification Model

The increased latency for correct rejection of repeated pseudowords observed in Exp.s 1 and 2 demands that repeated pseudowords be included in the verification set. To explain the increased latency for repeated pseudowords, the Verification Model must assume that the size of the verification set used with repeated pseudowords is larger than the size of the verification set used with novel pseudowords. Consider, pseudowords are identified by a failure to find a match in the verification set, and the time to verify a nonmatch is assumed to be a constant. Hence, it must be that the verification set size for repeated pseudowords is larger than that for novel pseudowords. The only event that will explain this increase in set size is the addition of a repeated pseudoword to the verification set, along with the assumption that the verification process fails to detect this potential match. The additional time for a non-match would lead to the increased time of rejection for repeated pseudowords.

The inclusion of repeated pseudowords in the verification set would also explain the increased false alarms seen for repeated pseudowords. However, if repeated pseudowords are to be included in the verification set, and the verification process fails to detect this match, then the verification process must be fallible. Furthermore, it must also be true that the verification process is less efficient at identifying late searched members. That is, the criteria for a match must be more stringent for terminal members of the verification set than for initial members of the verification set. Specifically, since repeated pseudowords would be late entries they are unlikely to be accepted as words. However, repeated pseudowords in both instruction groups show a similar 20 msec delay in response latency relative to novel pseudowords and, yet, there were more errors for repeated pseudowords in the speed condition. As a result, latency and accuracy in the Verification Model cannot be explained by a common process, they must be due to separate processes. While late verified words will show long latencies and high errors, the latency and accuracy observed must be due to a correlation between two different mechanisms and not to a common process. The interaction in errors for repeated pseudowords will also demand an independence between latency and accuracy, but more about that later.

The size of the Word Frequency Effect in Exp. 2 was virtually identical for both instruction conditions, 91 msec for accuracy and 89 msec for speed instructions. Yet the overall response latency for the two groups was quite different, 621 and 532 msec for the accuracy and speed conditions, respectively. That is, there was a main effect of word frequency and instructions in the latency of decisions. However, in response accuracy there was an interaction between word frequency and instructions. Compared to the accuracy condition, errors for rare words in the speed condition increased relative to errors for frequent words. As the Verification Model is a serial stage model then the effect of word frequency and the effect of task instructions must be located at different stages of processing. There are two stages in the Verification Model, the preprocessing stage and the verification stage, and the Word Frequency Effect is by definition located in the verification stage. Consequently, the effect of task instructions must have been located in the preprocessing stage. That is, verification with speed instructions must have begun sooner than verification with accuracy instructions. Moreover, because the average difference in latency between frequent and rare words was identical for the two groups, then the size of the verification set must also be identical for the two instruction conditions. The only other alternative, which assumes that set size differed and the average verification time for each item changed accordingly, is too much of a coincidence to be considered seriously.

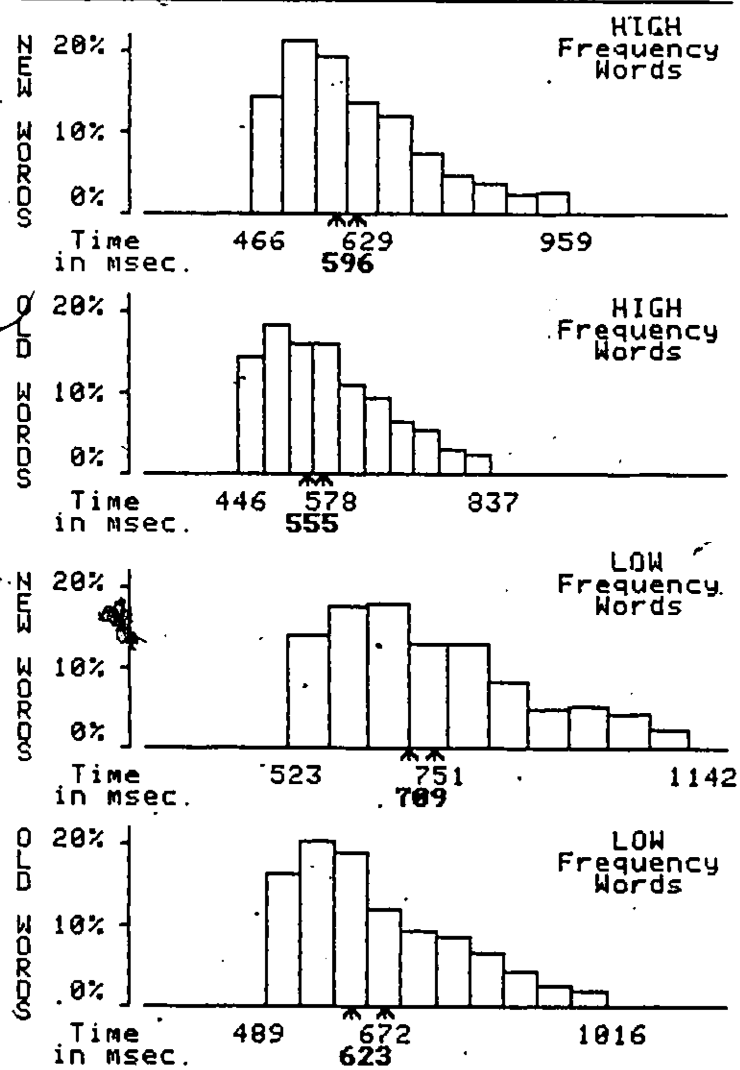
While the verification set was the same for speed and accuracy instructions, the two groups differed in accuracy. This reflects a difference in the quality of the sensory information used for decisions. However, while the quality of the information was different for the two conditions, it is not clear how the Verification Model would explain this difference. If a simple truncation notion were used to explain the increased errors to rare words, where the verification set is smaller in the speed condition, then the average verification time for rare words in the speed condition would be increased relative to the average time for frequent words. That is, there would be an interaction between instruction conditions and the Word Frequency Effect; yet there was no such interaction. The only alternative is to assume that late verified words in the verification set are verified less accurately and that late verified words are more susceptible to the decreased stimulus information accompanying early onset of verification in the speed condition. As a consequence, the Verification Model must assume that the latency and accuracy of a response are due to separate processes. This, of course, is the same conclusion as was used to explain the increased rejection latency of repeated pseudowords. In conclusion, while the Verification Model can account for the average word latency results in Exp. 2, the set of assumptions it must make, such as invariant verification set size across the two different levels of stimulus information and independence of latency and accuracy, reduces the plausibility of such a decision process.

Another objection to the Verification Model is the necessary assumption of an identical verification set even though the operation of the preprocessing stage has been drastically curtailed. This problem will be explored in more detail in the following section.

Latency Distributions. The distributions for frequent and rare words, and old and new words are shown in Figure 8 for accuracy instructions, and in Figure 9 for speed instructions. The mean minimum and mean maximum response times averaged over subjects are included, as well as the mean and median response times. The median response time is given in bold print. As can

69a

Fig. 8 Latency Distributions, for subjects in Exp. 2 receiving ACCURACY INSTRUCTIONS



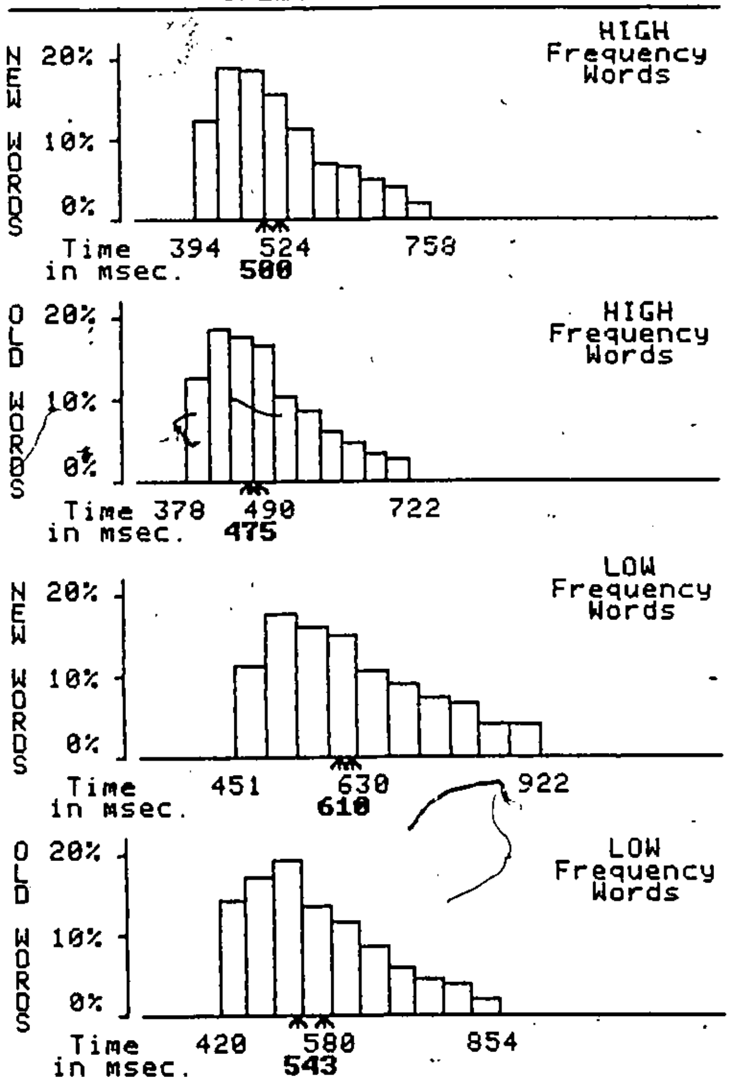
Where the latencies under the figures

are: The mean minimum latency
The mean response latency
The mean maximum latency

and The median response latency is in **BOLD** print.

The 'Y' axis contains the per cent responses in each of the TEN equal time bins.

Fig. 9 Latency Distributions, for subjects in Exp. 2 receiving SPEED INSTRUCTIONS



Where the latencies under the figures

are: The mean minimum latency
The mean response latency
The mean maximum latency

and The median response latency is in **BOLD** print.

The 'Y' axis contains the per cent responses in each of the TEN equal time bins.

be seen, all distributions show a similar, marked positive skew as was reported by Ratcliff (1978) for recognition times.

Of interest, there was a clear effect of instructions on the variance (as estimated by the range of latencies) of the latency distributions, with variance for the frequent words decreasing by 129 and 47 msec for new and old words (the variance of accuracy instructions minus the variance of the speed instructions), and variance for rare words decreasing by 148 and 93 msec, for an average of 105 msec. Note that the mean minimum response time decreased by approximately 70 msec in all four comparisons between the accuracy and speed instructions, while the mean maximum response time decreased by an average of 175 msec. The speed instructions shifted the latency distributions as a whole toward shorter latencies as well as reduced the variance associated with each condition. This pattern of decreased variance as a function of the mean response latency is not surprising for criterion bias models of lexical decision, such as the Logogen System or the Relative Familiarity Judgment model. Given the similar effects in latency found here for the effect of word frequency, the Verification Model would not initially predict the present correlation between the overall mean and variance. However, if it is assumed that response variance is a sum of two normal distributions, from the preprocessing and the verification stages, and that the variance and mean of the preprocessing stage decreased for the speed condition, then the Verification Model also can explain these distributions.

There is a cost to this assumption, though. As was discussed above, the Verification Model must assume similar verification set sizes in order to explain the similar effect of word frequency on decision latency in the two instruction conditions. Moreover, while the Verification Model assumes that the preprocessing stage establishes the verification set, it must now claim that similar verification sets are generated across large changes in the operating parameters of the preprocessing stage. This assumption of an invariance in output across large changes in the mean and variance is not a very plausible one given the observed changes in response accuracy. Thus, while the Verification Model may account for the average response latencies and latency distributions seen in Exp. 2, it can only do so by assuming that the output from the prepocessing stage is insensitive to its operating characteristics. Time dependent models of decisions such as the random walk process and the Logogen System can account for these correlated variances and means with fewer assumptions.

An interesting pattern in Figures 8 and 9 is that the mean latency for the frequent words in the accuracy condition and the mean latency for the rare words in the speed condition were virtually identical, 629 and 630 msec with new words, and 578 and 580 msec with old words, for the accuracy and speed conditions, respectively. This was surely coincidence! However, this does provide an opportunity to directly compare the latency distributions for frequent and rare words unconfounded by a difference in average latency. Not only are the means similar, the distributions are almost identical in all respects. Indeed on the basis of these distributions, it is tempting to conclude that the typical increase in variance observed with rare words is due solely to the greater average time spent in the system for rare words, and that except for this increased time to decision within an experiment, rare words and frequent words are subjected to similar time dependent decision processes.

<u>The Relative Familarity Judgment model</u>. The random walk diffusion process used by Link and Ratcliff is formally designed to deal with both match

and non-match processing and, unlike the Logogen System, it has separate processes for detection and rejection. Briefly, the Random Walk Diffusion Process models the accumulation of both feature matches and non-matches over time. Hence, feature processing is due to a time dependent similarity comparison between the target and memory information. When the stimulus evidence is decreased, the random walk process like the Logogen System predicts that relative to well known targets this decreased information will have a larger effect on the identification of poorly known targets.

Unlike the Logogen System, a similar but opposite result will occur for the distractors. Decreasing the stimulus evidence will have a larger effect on the rejection of more familiar distractors than on less familiar distractors. This result is diagramed in Figures 10a and 10b, which use data from the accuracy and speed conditions. Pseudoword distractors are shown on the left of the response criterion, while word targets are shown to the right of the criterion. Old words and new words here correspond to well known and poorly known targets, while old and new pseudowords correspond to familiar and unfamiliar distractors. As there is a single decision criterion for each of the tasks and as the four distributions are assumed to reflect the similarity between the targets and memory with similarity (or relatedness) increasing from left to right, then decisions should be most accurate for stimuli distant from the criterion (novel pseudowords and old words) and least accurate for stimuli close to the criterion (old pseudowords and new words) for both the speed and the accuracy condition. Ignore for now the labeling of the horizontal axis and assume that the distance from the criterion, in either direction, reflects increasing accuracy. Notice that as the stimulus evidence becomes less available, as in the speed condition, the average separation between pseudowords

Fig. 10a Normalized Accuracy Scores:
for ACCURACY INSTRUCTIONS in Exp. 2.
(from Table 9)

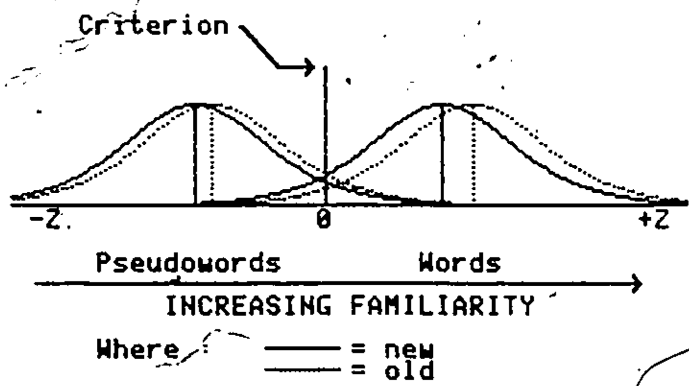
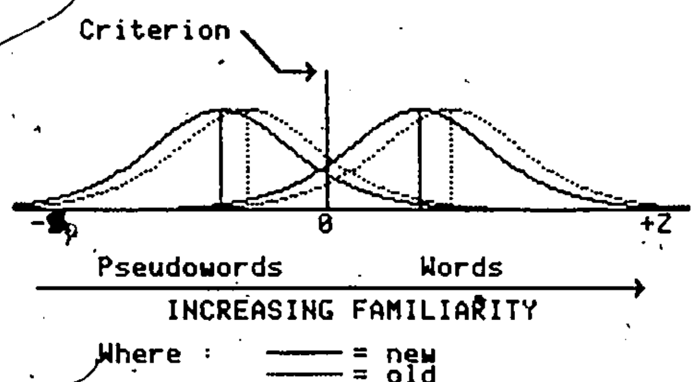


Fig. 10b Normalized Accuracy Scores:
for SPEED INSTRUCTIONS in Exp. 2.
(from Table 9)



and words decreases. As the old pseudowords and novel words are closer to the decision criterion, this decrease in stimulus evidence will lead to more errors for these items relative to new pseudowords and old words. That is, decreasing the stimulus evidence lowers the accuracy for low frequency words but increases the errors for repeated pseudowords. Notice also that the distance between the novel and the repeated pseudowords in Figures 10a and 10b are approximately the same. This constancy in separation between the normalized accuracy distributions for novel and repeated pseudowords implies that there was a similar increase in confusions for repeated pseudowords in the the speed and accuracy instruction conditions. Moreover, the relative difference in errors for repeated pseudowords between the two instruction conditions appears to be attributable solely to the overall level of performance in the two conditions, and not to a specific change in the processing of repeated pseudowords in the two instruction conditions.

The Effect of a Repetition

The discovery that word decisions are facilitated by a prior presentation (Scarborough, Cortese & Scarborough, 1977; Stanners, Jastrzembski & Westbrook, 1975) has been a major constraint for theories about the Lexical Decision Task. The explanations of this facilitation appeal to one of two mechanisms. Stanners et al. (1975) and more recently Becker (1976) have proposed that the facilitation is due to an active bias in the order of search through memory, while Scarborough et al. (1977) suggested that the benefit of a repetition is best interpreted as a change in some decision threshold, as in Morton's (1970) logogen theory, where the lowered decision threshold of a repeated word requires less evidence and hence less processing time. Despite their differences both theories locate the benefits of a prior presentation in a memory system that is specific to words. I would like to suggest an alternative explanation that is based upon a general characteristic of memory, and more specifically an explanation based on changes in grapheme familiarity. In this description of the repetition effect, a repetition does not influence word retrieval or lexical access but rather a repetition increases the information in memory about the characteristics of a particular stimulus. This change in memory information will usually facilitate decisions about words but hinder decisions about pseudowords.

Before continuing, it will be useful to review the results observed in Exp. 2 using a modified version of Signal Detection Theory. Table 9 presents the estimated <u>d′</u> and <u>Beta</u> for the novel and repeated words, using novel and repeated pseudowords as distractors. Notice that the effect of a prior presentation was to change the estimate of decision bias (<u>Beta</u>) and not the overall estimate of accuracy or precision (<u>d′</u>). In short, the facilitation often discussed in lexical decision may not necessarily be an improved abilty to locate a word in memory. Instead, it would appear to be a change in the willingness with which subjects accept a familiar (or repeated) word as a <u>word</u>, and may be independent of any formal lexical membership.

Figures 10a and 10b plot the normalized distributions using the previous empirical estimates of <u>d′</u> and <u>Beta</u> for <u>old</u> and <u>new</u> items, with word and pseudoword stimuli serving as target and distractor, respectively. These distributions can be viewed in terms of Signal Detection Theory by labeling pseudowords as <u>noise</u> and words as <u>signal+noise</u>. I have labeled the abscissa (the X-axis) as <u>increasing visual familiarity</u> for the purposes of this discussion. This, however, is a simplification as all factors that influence a decision contribute to performance on this axis; yet to the extent that

TABLE 9

Estimates of d′ and Beta in a Lexical Decision Task for Presentations and Instructions, in Experiment 2.

	TYPE OF INSTRUCTION			
	ACCURCY		SPEED	
	d′	Beta	d′	Beta
NEW	2.92	1.27	2.36	1.18
OLD	3.88	.52	2.42	.52

responses to pseudowords are mediated by some form of grapheme familiarity, this is a useful simplification.

I would like to draw attention to two aspects of Figures 10. First, notice that the effect of a prior presentation was to shift the curves for <u>old</u> items to the right of the curves for <u>new</u> items, and that this shift is identical in direction for both words and pseudowords. It would appear that a major portion of the benefit in a repeated word is independent of that word's abstract lexical status. This implies that repetition can influence graphemic processing as well as lexical processing. The second aspect of Figures 10, to which I wish to draw attention is the effect of a repetition with the speed and accuracy instructions. The effect of repetition with speed instructions appears to be more similar for words and pseudowords than for accuracy instructions. This may reflect different information used for lexical decision. That is, where performance in the speed instruction condition may reflect mainly visual information, and repetition would influence word and pseudoword information equally, performance in the accuracy instruction condition may reflect additional constraints in a decision, such as the availability of meaning, and repetition would would lead to different benefits for words and pseudowords. It is neither necessary nor appropriate to assume that all the benefits of a prior presentation reside at the graphemic level. As was discussed earlier, other properties, such as semantic or contextual information, may also be changed. However, the nature of this change and its benefits for performance will depend upon the actual task demands.

In summary, the simplest interpretation of the increased errors to repeated pseudowords and the increased accuracy to repeated words is to assume that subjects often treat the Lexical Decision Task as a general familiarity task (e.g., "Have I ever seen this string of letters before?"). As repeated pseudowords are likely to be more familiar than novel pseudowords, they are more likely to be incorrectly accepted as words. The proper question to ask about repetition, then, is what is the nature of the change of information in memory and not how are words accessed more quickly in some lexical memory.

In conclusion, it is proprosed that the response measured in lexical decision must, in part, be based upon the availability of a graphemic representation and that the memories measured by the Lexical Decision Task are composed of nonlexical as well as lexical attributes. An implication of this perspective is that we cannot conclude with any certainty that a word has been processed as a lexically meaningful unit when, and if, it shows a benefit of a prior visual presentation in the Lexical Decision Task. Conventional explanations of lexical decision ignore pseudoword performance, and hence are not easily modified to model changes for both words and pseudowords. Only an analysis of the Lexical Decision Task based on a common decision gradient for both words and pseudowords can easily account for the effects of repetition found in Exp.s 1 and 2.

Experiment 3

A major difference between the present account of the effects of repetition and that of the lexical store models is the degree of abstraction attributed to memory for words. The notion of a "dictionary-like" memory also assumes that words are represented at the highest level of abstraction, i.e., as a spelling pattern, and that details of a prior presentation such as typography and visual form are not maintained in this lexical memory. The relative judgment model does not assume that word knowledge is based upon the most general level of abstraction, as it is based upon an instance based structuring of memory (Brooks, 1978, 1982; Jacoby, 1982a, 1982b; McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982). Representations of words are not treated as central and summary descriptions of invariant letter relationships, but rather as multiple and distributed traces of the word. These traces maintain information about incidental visual detail associated with the encoding of the word, as well as information about inter-letter relationships. This memory for instances has implications for the current description of the repetition effect. Memory for a prior presentation should contain information about visual detail as well as information about the symbolic (letter) interpretation. The relative importance, however, of these different types of information in producing an effect of repetition can only be determined empirically and will depend upon both the type of information used at the time of test, as with semantic information in Exp. 2, and the relative availability of functionally similar resources. The latter constraint is needed to account for the different benefits of repetition seen for frequent and rare words, and reflects the fact that a prior presentation will be useful only to the extent that it represents an operationally significant change in memory information. In summary, the change in performance induced by a prior presentation is not necessarily limited to the repetition of symbolic information (i.e., a spelling pattern), repetition of other features such as visual detail may also be useful.

The following experiment was designed to investigate the nature of changes in memory for a graphemic stimulus that support a repetition effect in the Lexical Decision Task. The procedure involved a simple manipulation of upper and lower case information across a study and test session. If the benefit of a prior presentation is specific to the repetition of visual form (i.e., the shape of the word envelope and the type of typography), facilitation by repetition will be present only when a word is repeated in the same case. On the other hand, if facilitation by repetition is located at the level of graphemic structure or above, then the facilitation from a repetition should be independent of case variation. Mixed results are possible, of course, and in light of the previous discussion it was expected that there would be an interaction between case specificity and the type of stimulus. It was expected that repeated pseudowords would show greater specificty in visual form than repeated words and that repeated rare words would show greater specificity of visual form than repeated frequent words.

An additional factor investigated in this study was the effect of case information in lexical decisions. Words in lower case letters, potentially, offer more opportunities for distinctive word shapes (Garner, 1981). If words are processed in terms of holistic visual features, as well as in terms of letter relationships, then decisions about words seen in lower case, which contain more information, should be faster and more accurate than decisions about the same words when seen in upper case. The evidence for holistic processing with words is mixed. While some studies report no evidence of holistic processing (Scarborough et al., 1977) others have found some additional benefit of word shape in word processing (McClelland, 1976; Rayner & Posnansky, 1978). It would appear that at most holistic relationships only play an auxillary role in word identification. However, even an occasional contribution of holistic relationships invalidates the abstract word representations typically assumed by the lexical store theories. In contrast, the auxillary role of holistic relationships is compatible with an instance based theory of word memory. Thus evidence of a difference in the processing of upper and lower case words in the following experiment was taken as support for a multiple trace notion of word memory. Unfortunately, the converse, where there was no difference due to case, would not necessarity rule out holistic encoding, if this holistic encoding occurred at an abstract level.

An additional factor in the discussion of holistic relationships is the overall ease of processing. In keeping with the previous discussions of adding information, it might be expected that any difference in processing between upper and lower case words would be most noticable for rare words and least noticable for frequent words, and would be more noticable in first presentations than repetitions of words. Thus, as with repetition, the additional constraint provided by additional information is most noticable for poorly known words. It is not clear what predictions should be made for pseudowords. However, a general prediction is that case effects should be larger for familiar, that is repeated, pseudowords than novel pseudowords. As for words, variations in information should have a larger effect for difficult decisions than for easy decisions.

Method

Subjects. Thirty-two McMaster undergraduates were recruited from an introductory Psychology course, and received course credit for participating in a 45 min session.

Materials. The stimuli were the same 128 word and 128 pseudoword stimuli used in Exp. 2. The word stimuli contained four classes - the 2 x 2 factorial combination of High and Low Frequency Words, and concrete and abstract words. The pseudowords, also, contained four classes - the 2 X 2 factorial combination of 5 and 6 letter pseudowords, and monosyllabic and bisyllabic pseudowords. While the word stimuli contained the same structural classes as the pseudowords, the analysis for the words ignored these classes. An additional set of 25 words and 25 pseudowords were collected for practice items, 20 items for the letter search task and 30 items for the Lexical Decision Task.

Design. The stimuli were broken into 8 sets of 32 items, consisting of 16 words and 16 pseudowords. Each set contained 4 stimuli from each of the 4 word and 4 pseudoword classes. For any one subject, 4 stimulus sets were seen in both the Phase I and Phase II tasks, while the other 4 sets were seen only in the Phase II task. Consequently, for every 8 subjects, each stimulus was seen equally often as a once presented and a twice presented item. Further, each stimulus was seen in a 2 X 2 factorial combination of upper case and lower case typography during study and test. A complete couterbalancing of stimulus class, presentation, and study-test typography was used - a 4 X 2 X 4 factorial combination, which required all 32 subjects.

The Phase I, incidental learning or study task consisted of a letter search task, with 64 target-present (match), and 64 target-absent (non-match) trials. The match and non-match trials were counterbalanced appropriately for the 4 word and 4 pseudoword classes. Half the match and half the non-match trials were for target strings presented in upper case letters, while the remainder were for targets in lower case. Words and pseudowords were presented equally often in the two cases. In the match condition, each letter position in the target string was tested equally often. For all trials, the case of the target letter was the same as that for the following letter string. This consistency of case between the probe and target was used to encourage the use of a visual match strategy, rather than a name match strategy (see Posner, Boise, Eichelman & Tailor, 1969).

The Phase II test consisted of a Lexical Decision Task with 128 words and 128 pseudowords. Half the words and half the pseudowords were presented in upper case, while the remainder were presented in lower case. For half the repeated words and pseudowords, the stimuli were presented in the same typography in both Phase I and Phase II, while the remainder were presented in the alternate typography (i.e., of the stimuli presented in lower case in Phase I, half were presented in lower case in Phase II, while the other half were presented in upper case). The appropriate counterbalancing was carried out for the 4 classes of words and 4 classes of pseudowords for each subject, such that half the stimuli were presented in the same case during study and test, while the remainder were presented in the opposite case. Similar counterbalancing was carried out for the experimentally novel words and pseudowords.

Procedure. Each subject was tested individually on an APPLE-II plus computer, which controlled stimulus presentation and collected the type of response and the response latency to the nearest millisecond.

An experimental session began with 20 practice trials for the letter search task in which the subjects were allowed to familiarize themselves with the experimental equipment and procedure. The stimuli were presented as white letters on a black TV screen, and viewed at a distance of 1-2 feet. Each letter in the display subtended approximately .20 degrees horizontally and .25 degrees vertically. A responses required a key press to one of two telegraph keys interfaced with the computer.

Each letter search trial consisted of a 1 sec display of the probe letter, above and at the center of a following target string. Immediately preceding the onset of the probe letter, a soft auditory cue served to remind the subjects that a trial had begun. The target string was presented 750 msec following the removal of the target letter, and remained on the screen for 500 msec following the subject's response. This delay in erasing the target string allowed the subjects to monitor their own accuracy. The next trial began with the presentation of the next letter target, 1 sec after the subject's response to the previous target string.

Immediately, following the Phase I task, subjects were given a 5 min rest period, after which they were given instructions for the Phase II task. Immediately following these instructions, subjects were given 30 practice trials for the Lexical Decision Task. Subjects were instructed to respond as quickly as possible with a lexical decision, while maintaining accuracy. No feedback was given about the accuracy or speed of responses. The subjects were instructed, though, to monitor their own performance and if they found themselves making more than a few errors, they were instructed to take more time for a decision.

A lexical decision trial began with a soft auditory cue immediately preceding the onset of a visual warning signal. The visual cue, consisting of three plus signs, was presented for 750 msec and doubled as a fixation stimulus. The target string was presented in the same location as the warning stimulus, 500 msec after the warning signal was removed. The target string remained visible until the subject made a response, by pressing one of two response keys. The next trial began 1 sec following a response.

<u>Results</u>

The mean latency and accuracy was computed for each subject, for each of the stimulus conditions and entered into an analysis of variance. The effect of a prior repetition was assessed by (a) comparing the repeated stimuli with each other (<u>yes</u> and <u>no</u> responses were collapsed to increase the number of observations per cell); and (b) comparing the 4 word and 4 pseudoword classes collapsed across the Phase I conditions, with the appropriate novel stimulus. The analyses for words and pseudowords, accuracy and latency, were conducted separately.

Repeated Words. The mean latency and accuracy for the 16 conditions is given in Table 10. The latency and accuracy data were analyzed separately in terms of test case, same or different case as in phase I, word frequency, and meaning.

The repetition of words in the same or different case was a significant factor in only one comparison, a three way interaction involving word frequency and typography, F (1,31) = 6.08, p< .02. However, a subsequent analysis using study and test case as main effects showed this to be entirely due to an interaction of study case and frequency, F (1,31) = 6.08. There was no interaction between study and test case for any comparison involving study and test case, F < 1 for all interactions. As can be seen in Table 11, the interaction between study case and word frequency reflects different benefits of study case for frequent and rare words. Decisions for repeated and frequent words were faster when studied in upper case, while decisions for repeated and rare words were faster when studied in lower case. Response accuracy was inversely related to latency, hence there was no speed accuracy trade-off.

There was a significant effect of test typography in latency, F (1,31) = 16.54, p< .0003, with words when tested in lower case responded to faster than in upper case (583 vs 612 msec, respectively). Interestingly, there were two marginal interactions involving typography in response latency, one with word frequency, F (1,31) = 4.00, p< .06, and one with meaning, F (1,31) = 3.95, p< .06. In these comparisons, the difference between lower and upper case typography was largest with rare and abstract words, and smallest with frequent

TABLE 10

Mean Reaction Times (in msec)

and Percent Accuracy for Repeated Words

, in Experiment 3.

	HIGH FREQUENCY WORDS				LOW FREQUENCY WORDS			
MEANING:	CONCRETE		ABSTRACT		CONCRETE		ABSTRACT	
TEST CASE	RT	ACC.	RT	ACC.	RT	ACC.	RT	ACC.
SAME CASE								
:UPPER	543	(96.9)	557	(95.3)	647	(82.0)	707	(81.3)
:lower	548	(97.7)	552	(93.0)	617	(88.3)	616	(84.4)
DIFFERENT CASE								
:UPPER	566	(96.9)	569	(92.2)	624	(82.8)	687	(78.9)
:lower	531	(96.9)	536	(95.3)	620	(82.0)	650	(78.1)

TABLE 11

Mean Reaction Times (in msec) and Percent Accuracy
for the two way Latency Interaction involving:
Type of Study Case * Word Frequency,
for WORDS in Experiment 3.

		WORD FREQUENCY		
		LO	HI	L-H
UPPER CASE IN PHASE I.	RT	656	541	112
	ACC	82%	96%	
lower case IN PHASE I.	RT	636	559	77
	ACC	84%	95%	

Where: LO = LOW FREQUENCY, OR RARE WORD.
HI = HIGH FREQUENCY, OR FREQUENT WORD.

and concrete words.

Additional Results. There was a main effects of word frequency in both latency, F (1,31) = 89.3, p< .0001, and in accuracy, F (1,31) = 56.70, p< .0001, where frequent words were responded to faster and more accurately (550 msec & 95.5%) than were rare words (646 & 82.2%). Further, there was a main effect of meaning in latency, F (1,31) = 7.41, p< .01, and in accuracy, F (1,31) = 7.40, p< .01, where concrete words yielded faster and more accurate responses than abstract words. Finally, there was a two-way latency interaction involving word frequency and meaning, F (1,31) = 5.96, p< .02, where the effect of meaning was significant only for rare words (627 vs 665 msec, p< .01 by Scheffe's, for concrete and abstract words) but not for frequent words (547 vs 554 msec, respectively).

Repeated Pseudowords. Table 12 gives the latency and accuracy for the 16 conditions. The latency and accuracy of responses were analyzed for test case, same or different case as in Phase I, pseudoword length, and syllable length.

The repetition of pseudowords in the same or different case was a significant factor in only one comparison, a three way interaction involving typography and syllable length, F (1,31) = 14.70, p< .001, in latency. However, as for words, a subsequent analysis using study and test case as main effects showed this interaction to be based entirely upon the two way interaction of study case and syllable length, F (1,31) = 14.70. No comparison involving both study and test case approached significance. Table 13 presents the mean latency for each of the four conditions of the two way interaction. As can be seen the effect of syllable length reversed across study case. Bisyllable pseudowords took longer to reject than monosyllable pseudowords when pseudowords were

TABLE 12

Mean Reaction Times (in msec) and Percent Accuracy for Repeated Pseudowords, in Experiment 3.

TEST CASE	MONOSYLLABIC PSEUDOWORDS				BISYLLABIC PSEUDOWORDS			
	5-LETTER		6-LETTER		5-LETTER		6-LETTER	
	RT	ACC.	RT	ACC.	RT	ACC.	RT	ACC.
SAME CASE								
:UPPER	710	(83.6)	747	(82.0)	730	(82.8)	782	(74.2)
:lower	747	(76.6)	786	(71.1)	719	(82.0)	753	(66.4)
DIFFERENT CASE								
:UPPER	731	(86.7)	750	(76.6)	680	(87.5)	733	(75.0)
:lower	719	(71.1)	730	(75.8)	723	(83.6)	766	(66.4)

TABLE 13

Mean Reaction Times (in msec) and Percent Accuracy for the two way Latency Interaction involving: Type of Study Case * Syllable Length, for PSEUDOWORDS in Experiment 3.

		SYLLABLE LENGTH		
		MS	BS	M-B
UPPER CASE IN PHASE I.	RT	726	750	-24
	ACC	78%	77%	
lower case IN PHASE I.	RT	753	721	32
	ACC	78%	78%	

Where: MS = MONOSYLLABIC PSEUDOWORDS.

BS = BISYLLABIC PSEUDOWORDS.

studied in upper case, while monosyllabic pseudowords took longer to reject when pseudowords were studied in lower case.

The type of typography had a main effect on pseudoword accuracy, F (1,31) = 13.89, $p < .001$, with decisions about pseudowords in upper case showing greater accuracy than decisions to the same pseudowords when in lower case (81.1% vs 74.1%, respectively).

Additional Results. As found in Exp. 2, there was a main effect of pseudoword length in both decision latency, F (1,31) = 12.09, $p < .002$, and in decision accuracy, F (1,31) = 16.54, $p < .0003$, where 5-letter pseudowords were responded to faster and more accurately (720 msec & 81.7%) than 6-letter pseudowords (756 msec & 72.9%).

Old vs New: Words. Table 14 presents the mean latency and accuracy for the eight conditions.

There was a main effect of test typography in decision latency, F (1,31) = 11.58, $p < .002$, where words tested in lower case produced faster decisions than words tested in upper case (603 vs 628 msec, respectively); and an effect of test typography in decision accuracy, F (1,31) = 4.36, $p < .05$, where words judged in lower case produced greater accuracy than the same words when judged in upper case (87.6% vs 85.6%, respectively). In short, word decisions appeared to be fastest and most accurate when the words were seen in a lower case typography. Additionally, there was a two way latency interaction involving test case and meaning, F (1,31) = 5.30, $p < .03$, were differences due to meaning were significant only in upper case words, and the difference due to test case was significant only with abstract words ($p < .02$, by Scheffe's in both comparisons). Of further interest was a marginal three way interaction involving test case, word frequency, and meaning, F (1,31) = 3.53, $p < .07$, where

TABLE 14

Mean Reaction Times (in msec) and Percent Accuracy

for

OLD vs NEW Words, in Experiment 3.

	HIGH FREQUENCY WORDS				LOW FREQUENCY WORDS			
MEANING:	CONCRETE		ABSTRACT		CONCRETE		ABSTRACT	
CASE	RT	ACC.	RT	ACC.	RT	ACC.	RT	ACC.
NEW: UPPER	580	(93.8)	602	(92.6)	675	(74.6)	734	(70.7)
NEW: lower	561	(96.9)	576	(92.2)	667	(79.3)	688	(74.6)
OLD: UPPER	552	(96.6)	562	(93.8)	632	(82.4)	690	(80.1)
OLD: lower	537	(97.3)	544	(94.1)	616	(85.2)	632	(81.3)

the advantage of making decisions to words in lower case was largest in abstract and rare words, and smallest in concrete and frequent words. While this was not a significant interaction, it does suggest that the advantage of seeing a word in lower case becomes more pronounced as knowledge about a word is decreasing.

There was a main effect of repetition on latency, $\underline{F}$ (1,31) = 62.13, $\underline{p} < .0001$, and on accuracy, $\underline{F}$ (1,31) = 16.09, $\underline{p} < .0005$, with repeated words yielding faster RTs and greated accuracy (595 msec & 88.9%) than experimentally novel words (635 msec & 84.3%). Thus, the lack of case specificity observed for repeated words cannot be due simply to a ceiling level of performance; decisions to repeated words were facilitated by a prior presentation but this facilitation was indifferent to the repetition of words in the same or different typography.

Additional Results. As usual there was a main effect of word frequency in latency, $\underline{F}$ (1,31) = 168.87, $\underline{p} < .0001$, and in accuracy, $\underline{F}$ (1,31) = 108.39, $\underline{p} < .0001$. In addition, there was a two way interaction between word frequency and repetition in latency, $\underline{F}$ (1,31) = 4.36, $\underline{p} < .05$, and in accuracy, $\underline{F}$ (1,31) = 7.49, $\underline{p} < .01$), in which rare words were more difficult to judge, and benefited more from a prior presentation, than did frequent words.

Meaning yielded a main effect for latency of decision, $\underline{F}$ (1,31) = 18.77, $\underline{p} < .0001$, and accuracy, $\underline{F}$ (1,31) = 17.59, $\underline{p} < .0002$, and a two way interaction with word frequency for latency, $\underline{F}$ (1,31) = 7.36, $\underline{p} < .01$. As usual concrete words yielded faster and more accurate responses, than abstract words. There were no differences of meaning in frequent words.

Old vs New: Pseudowords. The mean latency and mean accuracy for the eight conditions are given in Table 15.

There was a main effect of test typography for accuracy, $\underline{F}$ (1,31) = 11.60, $\underline{p} < .002$, with decisions for pseudowords tested in upper case producing

TABLE 15

Mean Reaction Times (in msec) and Percent Accuracy for OLD vs.NEW Pseudowords, in Experiment 3.

TEST CASE	MONOSYLLABIC PSEUDOWORDS				BISYLLABIC PSEUDOWORDS			
	5-LETTER		6-LETTER		5-LETTER		6-LETTER	
	RT	ACC.	RT	ACC.	RT	ACC.	RT	ACC.
NEW:UPPER	702	(87.5)	712	(85.2)	700	(87.9)	736	(80.5)
NEW:lower	696	(83.2)	733	(79.7)	721	(84.4)	732	(82.0)
OLD:UPPER	717	(85.2)	749	(79.3)	702	(85.2)	752	(74.6)
OLD:lower	727	(73.8)	747	(73.4)	725	(82.8)	752	(66.4)

more accurate responses than the same pseudowords when tested in lower case (83.3% vs 78.2%, respectively). Note that the effect of test typography was reversed for pseudoword and word accuracy. More will be said about this shortly.

As found previously, there was a main effect of repetition in latency, $\underline{F}$ (1,31) = 6.31, $\underline{p}$< .02, and accuracy, $\underline{F}$ (1,31) = 25.57, $\underline{p}$< .0001, with repeated pseudowords yielding slower and less accurate responses (739 msec & 77.6%; respectively) than novel pseudowords (717 msec & 83.8%, respectively). Further, as with words, the failure to find an effect of repetition contingent upon the repetition of case information cannot be based upon floor effect as repeated pseudowords showed large differences in performance relative to novel pseudowords.

<u>Additional</u> <u>Results</u>. There was a main effect of pseudoword length in latency, $\underline{F}$ (1,31) = 12.90, $\underline{p}$< .001, with 5-letter pseudowords yielding faster responses than 6-letter pseudowords (716 vs 739 msec, respectively), and accuracy, $\underline{F}$ (1,31) = 28.74, $\underline{p}$< .0001. In addition, there was a two way interaction between pseudoword length and syllable length, $\underline{F}$ (1,31) = 9.94, $\underline{p}$< .005, where, while in the same direction, 6-letter pseudowords were significantly less accurate than 5-letter pseudowords for only bisyllabic pseudowords (75.8% vs 85%, respectively; $\underline{p}$< .01 by Scheffe's), and not for monosyllabic pseudowords (79.4% vs 82.4%, respectively).

There was a four way accuracy interaction involving typography, repetition, syllable length and pseudoword length, $\underline{F}$ (1,31) = 5.58, $\underline{p}$< .025, which was largely uninterpretable. The main effects discussed above were in general consistent across this interaction but variable, with only one reversal in 24 comparisons.

Discussion

The results provide a clear answer to what type of visual information is facilitated by a repetition, and that is information about graphemic structure. Although there was a large effect of a prior presentation in both decision latency and decision accuracy with both words and pseudowords, there were no sizable effects of repeating stimuli in the same typography at study and test. In summary, the benefits of repetition observed for words, and the costs of repetition observed for pseudowords were not contingent upon the repetition of visual form; rather, these benefits and costs were based upon a symbolic representation of information, such as graphemic information or information about letter sequence.

While there were two interactions involving same/different typography, the interpretation of these interactions does not change the above conclusions. In both these interactions an analysis in terms of the main effect of study case showed the variability to be due to interactions of study case with a third factor; there was no variance due to the interaction of study and test case. It was found that frequent words benefited most from a prior presentation in upper case, while rare words benefited most from a prior presentation in lower case. The size of this reversal was small, about 13 msec for both frequent and rare words, and was not significant in a post hoc comparison for either frequent or rare words. While this interaction could be spurious, a possible cause might lie in processing differences between frequent and rare words, and the different processing demands for upper and low case stimuli.

Also, there was an interaction of study case with syllabic structure in pseudowords. This too was a complete cross over. A prior presentation in upper case inhibited bisyllabic pseudoword rejection more than a prior presentation in lower case. Conversely, a prior presentaion in lower case inhibited monosyllabic pseudoword rejection more than a prior presentation in upper case. Described another way, the effect of syllabic structure in repeated pseudowords is reversed by the type of study case. That is, a prior presentation of pseudowords in upper case resulted in the typical finding of a longer rejection time for bisyllabic pseudowords (Interestingly, the previous studies employed only upper case letters), while a prior presentation in lower case resulted in the opposite finding of a longer rejection time for monosyllabic pseudowords. One possible explanation is that the processing of syllabic structure is responsive to the differences in information between lower and upper case typography. That is, appearance rather than letter information might be more important for monosyllabic pseudowords, and hence a prior presentation in lower case leads to a more integrated memory trace.

In conclusion, while there is no clear explanation of the interactions of study case in repeated stimuli, the very occurrence of effects of study case is suggestive of letter irrelevant information in the memory for words and pseudowords accessed by the Lexical Decision Task.

There were several striking effects of test case. In general it was found that words when in lower case were accepted as words more readily, for both latency and accuracy, than when they were in upper case. Moreover, this effect of case tended to be larger for experimentally novel, rare and abstract words. This implies that the two cases did not differ simply in their legibility, otherwise we would expect a main effect of case on decision latency. This issue will be discussed more fully in a following experiment. However, it can be shown that a manipulation of visual quality has additive effects on word frequency.

The above pattern of results is easily interpreted within the relative familiarity model of lexical decision. Consider, a typical encounter with a word is with a lower case word; consequently, it might be expected that the memories which support word judgments are more compatible with a lower case presentations than with upper case presentations. Moreover, the interaction of case information with the overall difficulty of making a word decision, as measured by word frequency, would indicate that visual form is an auxillary souce of familiarity, rather than a primary form. This conclusion is reinforced by the failure to find consistent and significant benefits of same vs different typography for repeated words.

Neither, the Logogen System nor the Verification Model can easily explain differences in the latency and accuracy of lexical decisions based upon typography. In lexical store models, word knowledge is represented at the highest level of abstraction, a level which does not retain information about the details of the stimulus display. Consequently, to explain the main effect of word typography, the basic description of word knowledge used by these models must be changed.

There was also a main effect of typography in pseudowords decisions, although only in response accuracy. Interestingly; the effect of typography was opposite in direction to that seen for words. Pseudowords tested in upper case were rejected more accurately than when tested in lower case. The latency difference, while not significant, showed a similar reversal, a pseudoword tested in upper case yielding slightly faster decisions than when tested in lower case.

The reversed effect of case information with words and pseudowords is consistent with the relative familiarity explanation of the Lexical Decision Task. In general, it is claimed that the information gained from lower case words is similar in kind to the information gained from a prior presentation. That is, it is a gain in relatedness or familiarity. This comparison between repetition and case information is easier seen in Table 16 which presents the empirical estimates of <u>d'</u> and <u>Beta</u> for each typographic case, separately. Figure 11 presents the same data graphically. As can be seen, the general discriminability of words from pseudowords (as measured by <u>d'</u>) does not differ for the two typographies. Indeed, where it does differ, upper case presentations yield slightly greater discrimination than lower case presentations.

The majority of the difference between upper and lower case items is captured in the empirical estimate of <u>Beta</u>. The pattern of <u>d'</u>s and <u>Betas</u> is identical to the pattern of <u>d'</u> and <u>Beta</u> for repetition reported in Exp. 2 and replicated in the present experiment (see Table 16). It would appear that lower case stimuli are in some sense special. In the relative familiarity model, lower case stimuli are special because our experince with words is typically with words in lower case script. Hence, memories about words are often memories about lower case words. To the extent that lexical decisions are based upon the familiarity of visual information, decisions should show a bias for lower case stimuli.

In the relative familiarity model, the different <u>Betas</u> and similar <u>d'</u>s are explained by assuming a single decision criterion and multiple target distributions located on a single decision continuum. The subject's task is not simply judging whether an item is or is not familiar, but rather he or she must

TABLE 16

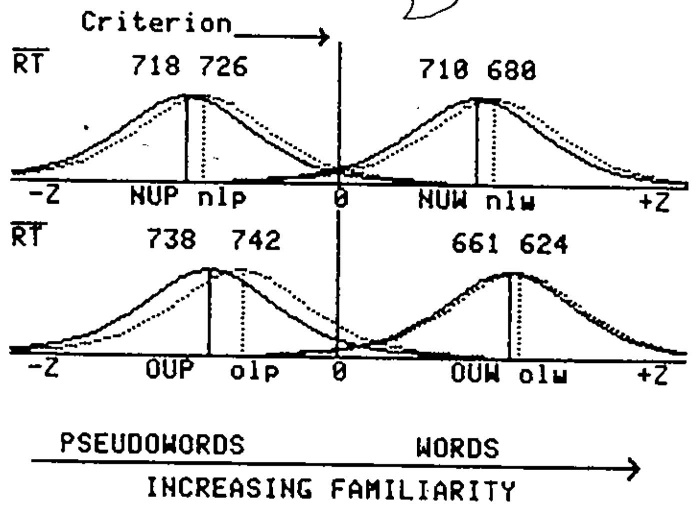
Estimates of d′ and Beta in a Lexical Decision Task for Typography and Presentations, in Experiment 3.

	TYPE OF VISUAL PRESENTATION			
	UPPER CASE		lower case	
	d′	Beta	d′	Beta
NEW	2.0	1.11	2.0	.87
OLD	2.1	.73	1.9	.56

Fig. 11 (from Table 16)

Normalized Accuracy Distributions:

for Change in TYPOGRAPHY, in Exp.3



Where: ——— = lower case words
——— = UPPER CASE WORDS

and: nlw = new, lower case words
NUW = NEW, UPPER CASE WORDS

olw = old, lower case words
OUW = OLD, UPPER CASE WORDS

nlp = new, lower case pseudowords
NUP = NEW, UPPER CASE PSEUDOWORDS

olp = old, lower case pseudowords
OUP = OLD, UPPER CASE PSEUDOWORDS

decide whether an item is more or less familiar than some criterion value of expected familiarity. The advantage of this analysis is that stimulus biases observed in performance and knowledge about those stimuli are linked. Multiple, systematic biases in decisions are expected when the test stimuli vary in how well they are known and when there is a single, common criterion for decision. As an example, the word bias for lower case items and the word bias for repeated items, shown in Table 16, is not necessarily due to two different decision biases, rather it may be attributed to systematic differences in the memory representations of lower case items, repeated items, upper case items and novel items relative to a single and common decision criterion (see Figure 11).

For general interest Figure 11 also contains the mean decision time above each distribution. Note that the mean latency decreases as the distance of the distribution from the decision criterion increases. As in Exp. 2, this demonstrates the typical correlation between the mean decision latency and the mean of the normalized accuracy distributions.

In summary, the results of the present experiment reveal that lexical decisions are influenced by the physical form of the stimulus presentation (e.g., upper or lower case), this was seen for both word and pseudoword decisions. However, the effect of a prior presentation appears to be largely independent of repeated "pictorial" properties and reflects instead memory for symbolic information, i.e., the graphemic information. Of interest, the difference in decisions to upper and lower case stimuli is reversed for words and pseudowords. This result implies a typography bias in lexical decisions. A typography bias is compatible with predictions on the basis of the Relative Familiarity Judgment model of lexical decision, but is incompatible with predictions from the Logogen System and the Verification Model.

Experiment 4

In Experiments 1, 2, and 3 similar patterns of results were found for word frequency and meaning. Both frequent words and concrete words were categorized faster and more accurately than were rare and abstract words. In addition, both the size of the Word Frequency Effect and the effect of meaning decreased with repeated words, in both decision latency and accuracy. While not likely, it was possible that the similar effects of word frequency and meaning were different levels of a single frequency dimension. That is, the effect of meaning was but a fine grained division of the Word Frequency Effect. Hence, within each frequency class concrete words would be more frequent than abstract words; this even though the concrete and abstract words were counterbalanced for word frequency in terms of the Thorndike - Lorge Norms (1944) of written word frequencies. In order to exclude this potential confounding of stimulus categories, a task was needed which would produce opposite effects for word frequency and meaning, or at the very least, a task which would affect one attribute and not the other. A possible candidate for this manipulation is seen in the reversal of the Word Frequency Effect typically found in episodic recognition (Glanzer & Bowles, 1976; Shepard, 1967). In episodic recognition, rare words are recognized more reliably than frequent words. Thus, the accuracy of responses is reversed for word frequency in recognition, as compared to lexical decision. This reversal can be put to use to test whether word frequency and meaning are related measures of word frequency. If recognition for concrete words is poorer than recognition for abstract words, then the hypothesis that meaning is a frequency measure is supported, whereas if recognition for concrete words is better than, or equal to, recognition for abstract words, this would support an independence between word frequency and meaning.

While there is no experimental evidence which deals specifically with concrete and abstract words, as operationally defined in the preceeding experiments, results from a study by Paivio & O'Neill (1970) on the effects of imagery in memory, suggested that concrete words should be recognized better than abstract words. As a consequence, in Exp. 4 it was hypothesized that in a test of episodic recognition, recognition for rare words would be better than recognition for frequent words, while recognition for concrete words would be better than recognition for abstract words. If these results were found, then the results would provide clear evidence of a functional independence between word frequency and meaning, when contrasted with the Lexical Decision Task.

Method

Subjects. Thirty-two McMaster undergraduates were recruited from an introductory psychology course, and received course credit for participating in a 45-min session.

Design. The stimulus materials and design were the same as that used in Exp. 3, with the exception that the Phase II task was changed to a recognition task. Thus the subjects were given a Phase II recognition test containing 128 old items and 128 new distractors (64 words and 64 pseudowords).

Procedure. The prodedure used in Exp. 3 was followed with two exceptions. First the Phase II practice trials were removed, and secondly the Phase II instructions were changed. Subjects were told to press one response key to indicate that yes they remembered seeing the item in the Phase I, letter search task, and another key to indicate no they did not recognize the test item.

Results and Discussion

Because of overall low level of recognition, the recognition scores were collapsed across case information. Examination of the means for accuracy revealed no differences due to case greater than 2%. The word data were analyzed separately for correct detections, and for false alarms, in a 2 X 2 design for word frequency and meaning. The mean percent correct detections and the mean percent false alarms are given in Table 17.

<u>Correct Detections</u>. Considering the hits for repeated words first, only word frequency yielded a significant effect, $\underline{F}$ (1,31) = 21.08, $\underline{p}$< .0001, with rare words being recognized better than frequent words (50.8% and 39.5%, respectively). The effect of meaning was not significant, $\underline{F}$ (1,31) = 1.35.

In summary, in episodic recognition, word frequency and meaning do not produce parallel effects. Indeed, as can be seen in Table 17, the ordering of the mean correct detections for the four conditions contradicts the assumption that meaning is a subdivision of word frequency. Specifically, recognition for abstract words, which to mimic rare words should have yielded higher recognition than concrete words, was lower than recognition for concrete words, although not significantly so. These results are all the more striking when it is considered that lexical decisions about rare, concrete words were almost intermediate between decisions to frequent, abstract words and decisions to rare, abstract words, for both response latency and accuracy. Consequently recognition for rare, concrete words should be intermediate to recognition for frequent, abstract words and recognition for rare, abstract words. Instead recognition for rare, concrete words was slightly better than recognition for rare, abstract words, while recognition for rare words in general was fully 13% better than recognition for frequent, abstract words. On the basis of this reversed effect for meaning and word frequency in episodic recognition, it is

TABLE 17

Correct Detections and False Alarms in Recognition, for Frequency and Meaning, in Experiment 4.

	HIGH FREQUENCY WORDS		LOW FREQUENCY WORDS	
MEANING:	CONCRETE	ABSTRACT	CONCRETE	ABSTRACT
CORRECT :	41.0 %	37.9 %	51.6 %	50.2 %
FALSE ALARMS :	22.7 %	23.2 %	20.1 %	30.8 %

possible to conclude that meaning and word frequency represent separate word attributes.

<u>False Alarms</u>. The analysis of the false alarms to new words produced a main effect of meaning, $\underline{F}$ (1,31) = 10.83, $\underline{p}$< .005, and an interaction between meaning and word frequency, $\underline{F}$ (1,31) = 6.80, $\underline{p}$< .015. As can be seen in Table 17, while the abstract words yielded more false alarms than the concrete words, this increase was significant for only the rare, abstract words ,$\underline{p}$< .01, by Schefe's.

The false alarm results support the previous conclusion that meaning and word frequency represent separate word dimensions. False alarms were greater for frequent words, when meaning was concrete; while false alarms were greater for rare words, when meaning was abstract. The interpretation of the large increase in false alarms for rare, abstract words is uncertain; however, it is possible that rare, abstract words are more often confused in terms of their meanings, than are rare, concrete words, or frequent words. Thus, in a semantically based check of the memory set, when rare, abstract words serve as distractors, they are more likely to overlap in meaning with a target word; hence they lead to more false recognitions. This account of the increased false alarms could be tested by presenting the test items in a semantically rich environment during the study trials, rather than in the semantically impoverished environment of a letter search task, as was used in the present experiment. This increase in semantic detail should decrease the false alarms to rare, abstract words relative to rare, concrete words.

In summary, the present results support the previous experimental division of meaning and word frequency, and imply that in lexical decision meaning and word frequency are not simply subsets of a common frequency dimension, but instead represent different aspects of a subject's word knowledge.

Pseudoword Responses. For completeness Table 13 presents the percent hits and false alarms for pseudowords. The responses were analyzed in terms of hits and false alarms for upper and lower case, for bysyllables and monosyllables, and for 5 and 6 letters. Correct hits at 40.6% were significantly greater than false alarms at 27.2%, F (1,31) = 60.11, $p < .0001$. There were no interactions for either hits or false alarms. Thus, while recognition for pseudowords was low, recognition of pseudowords was above chance, as the increased false alarms to repeated pseudowords in lexical decision would suggest.

Experiment 5

The discovery in Exp. 2 that a change in the decision criterion produces a main effect on response latency, and yet does not interact with the size of the Word Frequency Effect is similar to the finding that stimulus quality and word frequency have additive effects on response latency, reported by Stanners, Jastrzembski and Westbrook (1975) and Becker and Killion (1977). A change in stimulus quality for Stannners et al. (1975) entailed masking test stimuli with a random dot mask, and resulted in an equal increase in response latency for frequent and rare words. Becker et al. (1977) found a similar main effect of degradation when they decreased the visual contrast of the target stimuli. Indeed, it was on the basis of this result that Becker et al. (1977) argued against criterion bias model and for the Verification Model of lexical decision. Yet a similar additive effect in Exp. 2 was used for the opposite argument. The major factor in these different interpretations of similar latency patterns is the error data. While in Exp. 2 errors to high and low

TABLE 18

Correct Detections and False Alarms in Recognition, for Pseudowords, in Experiment 4.

	MONOSYLLABIC PSEUDOWORDS		BISYLLABIC PSEUDOWORDS	
	5-LETTER	6-LETTER	5-LETTER	6-LETTER
CORRECT :	33 %	45 %	41 %	43 %
FALSE ALARMS :	22 %	30 %	31 %	26 %

frequency words were found to interact with the task instructions, neither Stanners et al. (1975), nor Becker et al. (1977) reported this result for the manipulation of stimulus quality. While Becker et al. (1977) found a similar trend in one of their experiments, it was not significant. Surprisingly there was very little difference, in their reported accuracy, between frequent and rare words, and hence a possible interaction may be hidden by ceiling effects. Stanners et al. (1975) did not report their error data.

In view of the similarity between the manipulation of stimulus quality and of task instructions, and given the importance Becker et al. (1977) have attached to the additive effects in latency for word frequency and visual quality, it would be useful to determine whether there was also an error tradeoff for word frequency with a manipulation of stimulus quality. If the manipulation of stimulus quality and word frequency produce response patterns similar to those seen in Exp. 2, then much of the force of the arguments used by Becker et al. (1977) against the criterion bias models, and for the Verification Model, would be removed.

The following experiment was designed (a) to replicate the main effects of stimulus quality and word frequency, reported by Stanners et al. (1975) and Becker et al. (1977); and (b) to determine if, in contrast to the latency data, the decision accuracy for low frequency words was disrupted by degrading the stimulus quality more than decision accuracy for high frequency words.

An additional factor considered was the effect of degradation upon repeated pseudowords. On the basis of the arguments used in Exp. 2 to explain the influence of task instructions on pseudoword errors, it would be expected that 1) repeated pseudowords would produce more errors than novel pseudowords;

and 2) repeated and degraded pseudowords would yield more false alarms than repeated non-degraded pseudowords.

Method

Subjects. Twenty-four McMaster undergraduates were recruited from an introductory psychology course, and received course credit for this participation. Each subject served in one 45 min session.

Materials. The word and pseudoword stimuli were the same as those used in Exp. 1. There were 40 high frequency and 40 low frequency words, and 80 pseudowords. The stimuli were collapsed over the type of meaning and syllable length, and counterbalanced appropriately for these factors in the following design.

Stimuli were degraded by presenting words and pseudowords on a CRT behind a pattern mask. The mask consisted of a plastic transparency, cross-hatched by black square wave grids oriented at 45 and 135 degrees. The line width in the mask was approximately the same width as the letter stroke width for the CRT display. The mask served (1) to lower the stimulus intensity, and (2) to reduce the legibility of the display. While the stimuli when masked were difficult to read quickly, with a little effort they were easily identifiable.

Design. The 80 words and 80 pseudowords were randomly divided into two sets of 40 words and 40 pseudowords. These sets were counterbalanced across the other factors. Specifically, 4 conditions - the 2 X 2 factorial combination of type of stimulus display (non-degraded and degraded), and repeated or novel presentation during lexical decision. Thus, for every four subjects each stimulus was seen once in each of the four conditions.

Procedure. Subjects were run individually on a PDP-8A computer, which controlled stimulus presentation and recorded responses to the nearest millisecond.

An experimental session consisted of 30 practice trials, a Phase I study session and a Phase II test session. The practice, study and test sessions required subjects to make a lexical decision. Phase I served to pre-expose 40 words and 40 pseudowords, none of which were degraded. The sequencing of trials during practice and Phase I was identical to that used in Exp. 1. The Phase II test session presented half the words and pseudowords behind the mask (the degraded presentation), and the other half of the stimuli at normal contrast on the CRT (a normal presentation). Thus, for the repeated items, half were seen in Phase I and Phase II in a non-masked display, while the other half of the repeated words and pseudowords were seen masked in Phase II but had been seen in a non-masked display in Phase I. The sequencing of the test stimuli in Phase II was identical to that used in Phase I, with the exception that every second item was seen behind the mask. This was accomplished by presenting the non-masked stimuli above (about 3 degrees) the display position of the masked items.

The subjects were instructed to respond with a lexical decision as quickly and as accurately as possible, by pressing one of two response buttons. The word and the non-word buttons were counterbalanced across subjects.

<u>Results</u>

The mean latency and number correct were computed for each word and pseudoword condition, and used in an analysis of variance. Low frequency words only were analyzed for accuracy as there were only a few errors to high frequency words in the non-masked display condition. The mean latency and accuracy for the word stimuli are given in Table 19, while Table 20 presents

TABLE 19

Mean Reaction Times (in msec) and Percent Accuracy for Degraded and Non-Degraded Words, in Experiment 5.

	HIGH FREQUENCY WORDS				LOW FREQUENCY WORDS			
	NORMAL		DEGRADED		NORMAL		DEGRADED	
	RT	ACC.	RT	ACC.	RT	ACC.	RT	ACC.
NEW	553	(98.2)	665	(93.3)	644	(91.7)	737	(78.8)
OLD	512	(99.2)	616	(97.5)	590	(97.1)	690	(88.3)

mean latency and accuracy for the pseudowords.

Words. As expected degraded words took significantly longer to be judged relative to non-degraded words, $F\ (1,23) = 163.87$, $p < .0001$, 677 msec and 576 msec for degraded and non-degraded words, respectively. As well, there was an effect of degradation on accuracy, $F\ (1,23) = 24.04$, $p < .001$, with degraded low frequency words being recognised less accurately than non-degraded low frequency words (81% vs 93.3%, respectively).

As usual there was a main effect of word frequency on response latency, $F\ (1,23) = 132.00$, $p < .0001$, with high frequency words (536 msec) being responded to faster than low frequency words (677 msec), but frequency did not interact with degradation, $F\ (1,23) < 1.00$.

Unfortunately, the small number of errors for high frequency words made the planned comparison of accuracy between word frequency and visual quality impossible. Yet looking at Table 19, it can be seen that degradation decreased the accuracy of detection for low frequency words more than for high frequency words. While not strictly appropriate, an analysis of variance suggested that there was an interaction between word frequency and degradation, $F\ (1,23) = 7.57$, $p < .01$, using responses collapsed over repetition. Given the violated assumptions about variance, the estimated probability is likely incorrect, yet given the obtained significance level it is likely that the interaction was significant at a probability of at least less than .o5.

A better measure of the effects of degradation on the accuracy of decisions about frequent and rare words can be gathered from the estimated distance between the normalized high and low frequency distributions, which was .83 for the non-degraded words and .81 for the degraded words. Hence, the difference in normalized accuracy between high and low frequency words was normally additive (using the Z scores) across the manipulation of visual quality. As would be expected, analysis in terms of other distributions such as the Logistic Distribution revealed a similar main effect of degradation on the decision accuracy of frequent and rare words. As a consequence, it is clear that in terms of absolute number correct, decisions for rare words were hurt more by degradation than were decisions about frequent words. This response pattern where decisons about frequent and rare words show a main effect in response latency and an interaction in number correct is identical to that found in Exp. 2 with the exception that in Exp. 2 instructions manipulated the accuracy and speed of decisions while in Exp. 5 the visual quality of the stimuli was manipulated.

Additional Results. There was a significant main effect of repetition for latency, F (1,23) = 33.34, $p <$ .0001, and accuracy, F (1,23) = 16.94, $p <$ .001, with repeated words being both faster and more accurate than experimentally novel words. Repetition did not interact with degradation.

Pseudowords. Stimulus degradation slowed decisions for pseudowords, F (1,23) = 90.53, $p <$ .0001, with degraded pseudowords being responded to slower than non-degraded pseudowords (819 vs 698 msec, respectively). There was no significant effects of degradation on pseudoword accuracy. As usual, there was a significant main effect of repetition for pseudoword accuracy, F 1,23) = 8.43, $p <$ .01, with repeated pseudowords being identified less accurately than novel pseudowords (85% vs 91%, respectively). There were no other significant effects for either pseudoword latency or accuracy. The predicted interaction between repetition and degradation for pseudoword accuracy failed to reach significance, although as can be seen in Table 20 there seems to be a trend in this directon.

TABLE 20

Mean Reaction Times (in msec) and Perçent Accuracy for Degraded and Non-Degraded Pseudowords, in Experiment 5.

	TYPE OF VISUAL PRESENTATION			
	NORMAL		DEGRADED	
	RT	ACC.	RT	ACC.
NEW	690	(91.6)	825	(91.4)
OLD	683	(86.0)	797	(83.7)

Discussion

The results support previous experimental observations that word frequency and visual quality yield additive effects for the latency of a lexical decision. As will be remembered, it was this additivity in response latency that Becker and Killion (1977) used to argue against criterion bias models of lexical decision, and for the Verification Model. Becker and Killion (1977) claimed that as word frequency and visual quality are additive in response latency, they must act upon separate stages of word processing. Within the Verification Model this means that the absolute size of the verification set was the same for both degraded and non-degraded words, and that visual quality affected the time to generate the verification set and not the process of verification. Interestingly, given the pattern of errors in Exp. 5, the Verification Model must conclude that the act of verifying a word in the verification set does not guarantee that it will be detected. The verification set size for degraded and non-degraded presentations are assumed to be identical, and yet there were more errors for degraded rare words than for non-degraded rare words. Hence, as was also concluded in Exp. 2, the Verification Model must assume that the accuracy of a verification and the latency of verification are caused by independent processes.

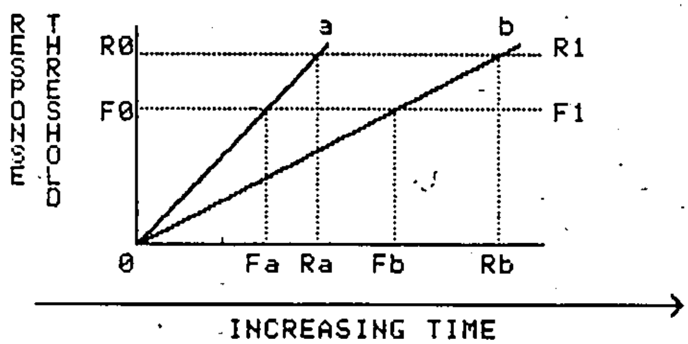
Criterion Bias Models

While the effect of degradation in Exp. 5 was not perfectly additive in latency, there was a 13 msec decrease in the Word Frequency Effect for decisions about degraded words, this difference did not approach significance, F (1,31) < 1; more important this difference was in the opposite direction to what Becker and Killion (1977) hypothesized for criterion bias models. They suggested that criterion bias models would predict an increase in the Word Frequency Effect as stimulus quality decreased. This prediction is presented graphically in Figure 12. In this figure, the ordinate indicates the amount of evidence necessary for a response, while the abscissa represents an increase in decision time. In this figure, the decision criterion for frequent words, line F0-F1) is lower than the decison criterion for rare words, line R0-R1; a criterion bias model assumes that a frequent word response requires less evidence than a rare word response. The two diagonal lines coming from the zero intercept represent the time dependent increase of information for two conditions that vary in overall stimulus quality, vector 0-a is for a non-degraded stimulus presentation, and 0-b is for a degraded stimulus presentation. As can be seen, both vectors cross the decision criterion for frequent words before crossing the decision criterion for rare words. Moreover, it can be seen that the increase in latency for degraded rare words, Rb - Ra, is larger than the increase in latency for degraded frequent words, Fb - Fa. On the basis of a similar analogy, Becker and Killion (1977) concluded that a reduced rate of information extraction in criterion bias models should interact with word frequency. Furthermore, as the latency responses for degraded words did not show an interaction across word frequency, Becker and Killion (1977) concluded that a criterion bias model cannot account for response latency in the Lexical Decision Task, and they stressed the Verification Model's ability to account for the main effects of visual quality and word frequency.

In Exp.s 5 and 2, however, it was found that while the Verification Model could account for the latency differences of decisions to frequent and rare words, it could not account for the differences in accuracy for these decisions. The differences in accuracy, on the other hand, are in general compatible with criterion bias models. In short, neither the Verification Model

Fig. 12

Time Dependent processing for CRITERION BIAS MODELS of Decision.



Where (0a) represents the rate of encoding for NON-DEGRADED WORDS,

and (0b) represents the rate of encoding for DEGRADED WORDS.

R0-R1 represents the response threshold for RARE WORDS.

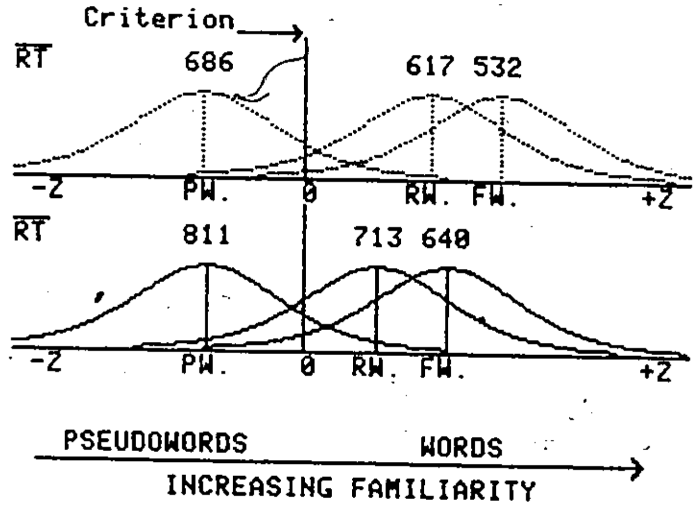
F0-F1 represents the response threshold for FREQUENT WORDS.

nor the criterion bias models can claim to be the best description of the Lexical Decision Task. Each type of model performs reasonably well on its chosen response variable, and poorly on the other. The following section discusses a possible modification of the time dependent accumulation of information used by criterion bias models, and shows how criterion bias models with this change can describe both the latency and accuracy of responding across manipulations of visual quality, and of accuracy instructions.

In its simplest form, criterion bias models predict a non-additive effect of stimulus quality and word frequency on response latency. However, if it is assumed that information is accumulated non-linearly in time, rather than linearly as in Figure 12, then the criterion bias models can be shown to predict additive effects between word frequency and visual quality. Note that this additivity in reponse latency is based on a similar additivity in response accuracy. It is proposed that the mean response latency should reflect the estimated mean of the normalized accuracy distribution. In order to demonstrate this relationship, Figure 13 portrays the normalized accuracy distributions for frequent words, rare words, and pseudowords as observed in the degraded and the non-degraded presentations. Following the previous convention, pseudowords are displayed to the left of a decision criterion, while words are shown to the right of the criterion. The mean correct response time for each type of stimulus is given above each distribution. As can be seen, the effect of a degraded presentation was to shift the mean of the distributions for both frequent and rare words equally toward a criterion. Hence, there was a simple and additive main effect of word frequency and visual quality in the normalized accuracy data, as well as in the latency data. If it is assumed that the average response latency and average normalized accuracy reflect a common

105a

Fig. 13 Normalized Accuracy Scores: for Change in VISUAL QUALITY IN Exp. 5



Where: ······ = NON DEGRADED WORDS.
——— = DEGRADED WORDS.

and: FW. = FREQUENT WORDS, or High Frequency Words.

RW. = RARE WORDS, or Low Frequency Words.

PW. = PSEUDOWORDS, or Non-Words.

process, then the additive effects observed in response latency are compatible with criterion bias models. Other non-linear transformations such as the logistic distribution also yield a similar prediction of additivity for response accuracy and latency.

The type of changes that need to be made in the random walk decision process to model a non-linear accumulation of information over time will be discussed in the final discussion. At present consider that if it is assumed that both decision accuracy and latency are in some manner similar to processes described by Thurston in his Law of Comparative Judgment (cited in Kling & Riggs, 1971), then it is possible to view both the accuracy and the latency of responding as related measures of a common process of decision, rather than as independent measures as must be claimed in the Verification Model.

While it is difficult to see how the Verification Model, given its assumptions of serial processing, can be modified to support comparative judgment, both the Logogen System and the Relative Familiarity Judgment model being statistical in nature can with only minor changes in assumptions make this shift.

In summary, the assumptions underlying the Relative Familiarity Judgment model of lexical decision can allow a simple description of the observed correlation between latency and accuracy of responses. While this descriptive power is not unique to the current theory, and most criterion bias models could be similarly modified, the advantage of the present model is that judgments about words and pseudowords can be considered as decisions based upon a common discrimination space in memory, the relatedness dimension and a common decision process, modeled by a random walk process. Other models of lexical decision, such as the Verification Model or the Logogen System, do not possess these advantages.

GENERAL DISCUSSION

Results of these experiments support the hypothesis of multiple information bases in the Lexical Decision Task. According to this hypothesis lexical decisions are based upon a relative weighing of several stimulus attributes, in accordance with task demands. The key predictions derived were that decisions would reflect both visual familiarity and semantic associations.

The role of visual familiarity was confirmed by the presence of increased errors to repeated pseudowords in Exp.s 1, 2, 3, and 5, and by the increased latency of correct rejections for repeated pseudowords in Exp.s 1, 2, and 3. Furthermore, finding that lexical decisions can be systematically biased towards lower case stimuli confirms the presence of visually based information in a lexical decision. Importantly, the interaction of the case bias with stimulus class, frequent or rare words, concrete or abstract words and old or new stimuli, implies that the case bias was not a simple response bias. That is, a bias in the absence of stimulus information. Rather, the bias interacted with the stimulus category. Highly familiar stimuli, frequent, concrete words showed little influence of the bias as did highly unfamiliar stimuli, novel pseudowords. It was stimuli of intermediate familiarity, rare, abstract words and repeated pseudowords, that gave the largest evidence of a case contingent decision bias. One explanation of this stimulus dependent bias would be to assume that case information entered into a decision only when subjects were guessing. As subjects would need to guess primarily for stimuli of intermediate familiarity then case information will affect mainly these lexical decisions. While I object to the introduction of a separate guessing stage, the implications of this perspective are identical to those proposed previously for the relative judgement hypothesis; namely that lexical decisions are multidimensional and reflect different sources of information.

The importance of semantic information for lexical decisions was supported by the consistent differences found for words which differed only in the meaning of their referent. The results showed a large and reliable benefit, in both decision latency and accuracy, for decisions about words with a concrete referent over words with abstract referents. Furthermore, in Exp. 2 it was found that a prior word presentation in a context that maximized semantic processing facilitated lexical decisions more than a prior word presentation in a context that minimized semantic processing. That is, there was facilitation from a prior semantic context. However, this facilitation of a prior semantic context was not an automatic consequence of the prior presentation. Subjects instructed to respond quickly, showed no facilitation due to prior semantic context; although decisions for previously seen words were facilitated. It was only when subjects were forced to respond accurately that a recent experience with semantic context facilitated a lexical decision. Presumably, the emphasis on accuracy emphasized a semantic verification of the visual familiarity information. This semantic check not only explains the increased benefit of a prior semantic context in word decision, it also accounts for the decreased errors made for repeated and hence familiar pseudowords. Indeed, the reciprocal benefits of enhanced pseudoword rejection and increased sensitivity to prior semantic context when subjects were given accuracy instructions makes a strong case for flexible criteria within the Lexical Decision Task. Clearly, when subjects are instructed to respond as quickly as possible with a lexical decision, the standards they employ leave the subjects vulnerable to vocabulary irrelevant experience, as is shown by the increased errors to repeated pseudowords, and insensitive to the details of a prior experience, as is shown by the insignificant effect of a prior semantic context. When subjects are asked to respond as accurately as possible with a lexical decision, decisions match vocabulary restrictions more consistently as is shown by the reduced false alrams to repeated pseudowords. However, this increase in selectivity is gained only at the cost of an increase in the attention paid to the semantic consequences of the familiar stimuli. Potentially, it is only by shifting the criteria from 'Does it look like a word?' using visual familiarity, to 'Does it function as a word?' using semantic context, that subjects can optimize vocabulary specific, lexical decisions. However, this shift in emphasis is only relative. When subjects were given accuracy instructions, it still took longer to reject repeated pseudowords than novel pseudowords. Consequently, subjects must still have been responding to visual familiarity, and it is possible that many highly familiar words and many highly unfamiliar pseudowords were responded to on the basis of this visual information.

In conclusion, the Lexical Decision Task does not measure a single and informationally restricted structure of memory, i.e., a Lexical Memory. Rather decisions are based upon the similarity of the present encoding with some set of acceptable alternatives in memory. This memory set is not necessarily restricted to vocabuluary members, or to semantically useful members, rather it is restricted to features correlated with these general constraints. However, not surprisingly a visual lexical decision often reflects visual familiarity, although it is the relative familiarity of the encoded stimulus that is important, and not some context free estimate of stimulus familiarity. In addition, decisions can be expected to reflect the overall distribution of targets and distractors as presented by the experimenter, and the task demands for accuracy and speed. The extent to which lexical decisions are based upon a semantic access to memories is unclear. In general, semantic constraints can be expected to increase in importance 1) as the target words decrease in familiarity, 2) as the pseudoword distractors increase in familiarity, and 3) as accuracy is emphasized. In conclusion, while the Lexical Decision Task measures word related knowledge, the type of knowledge it addresses will depend in part upon the test environment and the task demands. Hence, while the Lexical Decision Task may serve as a useful measure of word related knowledge, it cannot guarantee that its experimental measures are directly relevant to reading skills.

What is Facilitated by a Repetition?

Following Exp. 2 it was claimed that the repetition effect typically found in the Lexical Decision Task was best explained as an increase in grapheme familiarity, rather than a change in word availability. In large part this claim was based upon the contrasting effects of repetition for words and pseudowords. Repetition faciliated word acceptance but impaired pseudoword rejection. Obviously this claim rests upon the present characterization of the Lexical Decision Task, and has very clear limitations. A situation is easily imagined where familiarity would enhance pseudoword decisions. In particular, consider the case where the pseudowords are well learned, perhaps a list of nonsense strings learned to a 95% criterion in a serial list learning experiment. In this situation subjects might be expected to reject the learned pseudowords faster and more accurately than novel pseudowords. Clearly, familiarity is not necessarily bad for pseudoword rejection, indeed familiarity here would facilitate decisions by providing contextual information, such as

'This was in the memory list and hence it must be a nonsense word.'. What is important is the type of information stored about the pseudowords and the type of information used in a lexical decision. When a large proportion of the pseudowords are familiar, visual familiarity is no longer a useful dimension for lexical decisions.

An interesting example of response facilitation with pseudowords comes from experiments reported by Scarborough, Cortese and Scarborough (1977) and Kirsner and Smith (1974). The procedure in both of these experiments required subjects to make lexical decisions about a series of words and pseudowords in which some words and pseudowords were repeated following a fixed number of intervening decisions. A repetition could occur after zero intervening decisions (i.e., immediately), or after 1, 3, 7, 15 or 63 intervening decisions. For both experiments a large positive benefit in decision latency was reported for both pseudowords and words at levels of 0, 1 and 3. In contrast, with a larger number of intervening decisions, only word decisions showed a significant facilitation. Decisions for repeated pseudowords at larger intervals were not significantly faster than decisions for novel pseudowords. That is, decisions to words were facilitated by a prior presentation at all test levels, while decisions to pseudowords were facilitated most with an immeditate repetition and decreased as the number of intervening decisions increased. Indeed, with 63 intervening items, repeated pseudowords took slightly but not significantly longer to reject than novel pseudowords.

Increased familiarity does not necessarily hinder pseudoword decisions. In the above experiments, immediate repetition of a pseudoword not only provided information about its familiarity, it also provided information about the item and how it should be responded to. Hence, highly familiar pseudowords may provide contextual, or response, information as well as information about visual familiarity. Not surprisingly the positive benefits from contextual information may offset the negative effects of increased visual familiarity. If it is assumed that the availability of contextual information decreases more rapidly than the availability of visual information, then the positive facilitation for repeated pseudowords would be expected to decrease over time, and that eventually a prior presentation would show negative benefits. In Exp.s 1, 2, 3, and 5, the time between the first and second presentation of a pseudoword varied from 15-25 minutes, which was from 10-20 minutes longer than the longest interval used by Scarborough et al. (1977) or Kirsner et al. (1974). Also, given the decrease in facilitation with increasing intervals reported by Scarborough et al. (1977) and Kirsner et al. (1974), and given the negative benefits found for repetition of pseudowords in the previous experiments, it can be conclued that the effective value of pseudoword information changes across time. In short, whereas Scarborough et al. (1977) suggest that the activation from a prior pseudoword presentation decayed across time while the activation from a prior word presentation was maintained, a better explanation is to assume mixed costs and benefits for a pseudoword presentation, and that the usefulness of this information changes across time, while only benefits occur for words. This serves to mask any change in information supporting facilitated word decisions.

The experiment reported by Kirsner and Smith (1974) is of additional interest because they test the effect of pseudoword and word repetition across modalities as well as within modalities. Training and test could be in either an auditory and/or a visual modality. Interestingly, the facilitation found for repeated pseudowords at short temporal delays was restricted to decisions within the same modality (visual-visual, and auditory-auditory). Decision latency for pseudowords experienced in one modality and retested in the other modality (auditory-visual, and visual-auditory) were unaffected by a prior presentation. This finding supports the current argument that a pseudoword encoding is visually based for visual lexical decisions. On the basis of their data Kirsner and Smith (1974) suggested that the modality specific component in the Lexical Decision Task may be attributed to persistence in the non-lexical component of categorization. Their conclusion is similar to the present argument for a graphemic representation. However, results from Kirsner and Smith (1974) showed a similar repetition facilitation for words and pseudowords mainly at short intervals, and it was subsequently argued by Scarborough et al. (1977) that facilitation persists for words and decays for pseudowords. On this basis, Scarborough et al. (1977) argued for separate word and pseudoword mechanisms. Words were to be represented by enduring structures, such as a logogen, while pseudowords were only peripherally and temporarily represented. However, on the basis of the present data, it can be seen that pseudowords also leave enduring traces. The present explanation claims that the information contributing to a repetition effect changes across time, for both words and pseudowords; whereas for word decisions all information supports a repetition benefit, for pseudoword decisions only information available at short delays supports benefits. Hence, at longer delays only word decisions are facilitated by a prior presentation, while pseudoword decisions are inhibited. In summary, while Scarborough et al. (1977) wish to claim an unique status for a word unit on the basis of enduring effects of repetition for words but not pseudowords, the current results do not support a special status for word representations. Similar changes due to a prior presentation can be seen for words and pseudowords when the type of information and the criteria used for lexical decision are considered. It is the manner in which information influences a lexical decision that separates words from pseudowords, not the nature of the underlying representation, or the change in this representation.

In conclusion, a change in visual familiarity will impair pseudowords judgements only to the extent that decisions are based upon visual attributes. When pseudowords become highly familiar, contextual information is also available and will mediate Lexical Decisons. To the extent that pseudowords are defined to be word-like but visually unfamiliar stimuli then visual characteristics can mediate pseudoword decisions. This explanation is different from previous explanations, such as the Logogen System and the Verification Model, in that pseudowords are actively rejected on the basis of incomplete similarity information and are not simply rejected due to a failure in finding a lexical match.

Models of Lexical Decision

The next section will summarize the relevance of the present results for the three models of the Lexical Decision Task presented in the introduction. Following this there will be a more detailed discussion of how latency and accuracy are linked within Relative Familiarity Judgement; and a final section will discuss how experimenter controlled changes in task demands can lead to changes in the relatedness continuum supporting decisions.

The Verification Model. Several results of the current experiments are not consistent with predictions made from the Verification Model. These include the effect of semantic information in lexical decision, the increased latency and errors found for repeated pseudowords and the effect of case information in the latency and accuracy of word decisions. Moreover, while the Verification Model's explanation of the Word Frequency Effect is consistent with the latency results, it must assume that decision latency and decision accuracy are due to independent processes. This presents a more serious problem for the Verification Model as this implies that the size of the verification set would have to be independent of the quality of the stimulus evidence. That is, the verification set size would have to be a constant memory set rather than stimulus defined search set. This is a major modification.

Verification in the Verification Model is assumed to be based on a set of alternatives created by a preprocessing system similar to the Logogen System. The verification set by definition is a subset of lexical memory, and is generated on the basis of stimulus information by the preprocessing stage. While this initial stage extracts sufficient features to limit the number of lexical alternatives that have to be considered, processing at this level cannot select a particular word. Moreover, this passive preprocessing system is assumed to accumulate featural information over time and is influenced by stimulus quality (Becker & Killion, 1977).

It was found in Exp. 2 that for the Verification Model to be true, the time at which verification is begun must in some manner be under subject or task control. Latencies in the speed condition were 20% faster than latency in the accuracy condition, but the intensity and quality of the visual information was identical for the two tasks. Furthermore, as the effect of word frequency on latency was the same in both conditions, the verification process would have to be virtually identical for the two conditions. Hence the decreased response times must have been due to an earlier onset of verification when speeded decisions were required. It is problematic for the Verification Model that this early onset of verification had no effect on the verification set size, yet errors increased as latency decreased. Given the different levels of accuracy, it must be concluded that the quality of the stimulus evidence was different for the two tasks, and yet the verification set size as determined by the effect of word frequency on response latency did not reflect this change in stimulus evidence. As a result it appears that the set of alternatives used in verification must be independent of the quality of the stimulus information. A similar conclusion may be drawn from Exp. 5 where visually degrading the stimuli had a large effect on accuracy, but evidently no effect on the verification set size, as inferred by the similar Word Frequency Effect in decision latency for degraded and non-degraded stimuli. This observation that the effect of word frequency on decision latency is invariant across different levels of stimulus information reduces the plausibility of a verification process, in as much as it suggests that the verification set is not a stimulus relevant set of alternatives, but is a constant memory set. On intuitive grounds a serial search through all words in memory makes little sense. Indeed it was the limited number of word alternatives produced in the preprocessing of stimuli that gave the Verification Model plausibility as a psychological model.

In conclusion, the Verification Model explains the effect of word frequency on decision latency by assuming a biased serial verification of a stimulus-restricted set of altenatives. Yet it was found in Exp.s 2 and 5 that the Word Frequency Effect was constant across large changes in stimulus information. It is proposed that unless this insensitivity of the verification set size to stimulus information can be explained, then a fundamental assumption of the Verification Model is in error. There is no evidence for a preprocessing stage which defines the set of altenatives used in a subsequent verification stage. As a consequence, the Verification Model is not an adequate model of the Lexical Decision Task for either accuracy or latency.

The Logogen System. While the Logogen System can account for some of the present results, such as the effect of semantic information, and the main effect in normalized accuracy scores of word frequency with degradation and speed instructions, it is incompatible with other results. First of all, decisions cannot be localized as occurring at either visual input logogens or the Semantic System. The increased errors observed for repeated pseudowords requires that decisions occur at the level of visual input, while semantic information necessarily involves decisions in the Semantic System. Any attempt to localize lexical decisions at a specific and informationally limited representational system must be modified to accomodate the multiple constraints found for lexical decisions. Moreover, the present data for pseudowords suggests that the attempt to justify lexical decisions solely by appeal to vocabulary knowledge or meaning is too restrictive. While such vocabulary restriction may occur in elaborately processed stimuli, it is neither necessary nor, for well known stimuli, practical.

The discovery in Exp. 3 that lexical decisions can respond to information about visual form (i.e., case information) is not directly consistent with the Logogen System, although suitable modifications could be made in the visual input logogens. To do so, however, the logogens must either be increased in complexity to include case information, or two separate input systems must be created, one for each case. Regardless, the nature of a logogen will be changed from a general and abstract word detector to a general but restricted pattern detector.

The finding that errors to repeated pseudowords increased when speeded decisions were required in Exp. 2 is not only unpredicted by the Logogen System, it is the exact opposite of what the Logogen System would predict. Indeed any model of lexical decision which assumes that pseudowords are rejected by default (That is, pseudowords are identified by the failure to find a word match) cannot account for this phenomena. The discovery that sensitivity to familiar but poorly known distractors (i.e., repeated pseudowords) increases while the sensitivity to familiar but poorly known targets (i.e., rare words) decreases must be interpreted in terms of decisions to small differences on a common gradient. The results cannot be explained simply as an increase in the rate of guessing, as the normalized distance (i.e., <u>d'</u>) between frequent and rare words, and novel and repeated pseudowords were virtually identical for the accuracy and speed conditions (see Figures 10a and 10b). As a consequence an attempt to explain the differences in performance between the two conditions by appeal to a higher rate of guessing in the speed condtion must allow for more guessing with repeated pseudowords and rare words than for novel pseudowords and frequent words. In sum, it was not that information was absent in decisions about rare words and repeated pseudowords in either the speed or accuracy instruction conditions, rather the information present in the speed condition was simply not sufficient to reliably judge the stimuli relative to a criterion, whereas in the accuracy condition similar familiarity information was reliably discriminated from true <u>words</u>. It would seem that limiting the information in an encoding does not necessarily decrease the probability of finding the item in memory, it may simply decrease the probability of reliably discriminating the current stimulus with information in memory. The implications of this observation are important for all psychological theories about memory retrieval, as well as the Lexial Decision Task. For example, conventional theories of recognition memory assume that errors occur in recognition because subjects have difficulty finding a match between the test item and a prior episodic trace. An alternative, suggested by the above discussion, is that subjects have difficulty in discriminating a match, not in finding a match. Interestingly, if a memory task could make use of 'stored' constraints without necessarily isolating the source of these constraints from similar, and hence confusable memories, the effect of a prior presentation would be more robust than is generally allowed for by tests of memory such as recognition and recall, which assess temporal-contextual associations. A similar contrast of unique and shared information may operate in the dissociation of memory information across types of memory tests discussed by Jacoby and Dallas (1981) and Jacoby and Witherspoon (1982), while the process model of performance developed by McClelland and Rumelhart (1981) and Rumelhart and McClelland (1982), which is based upon memory constraints derived from an incomplete access to individual instances, may address only the retrieval of shared information.

In conclusion, the present data cannot be explained by any system which consists only of a set of match detectors varying in threshold, such as the Logogen System. Decisions are made on the basis of partial or incomplete information, and opposite effects are observed for targets and distractors as the average level of information changes. This implies a difficulty in discriminating a target, and not a failure to process or identify the target.

The Relative Familiarity Hypothesis. Relative Familiarity Judgment provides a clear description of the effects of repetition for latency and accuracy with both words and pseudowords and predicts opposite effects of a prior presentation when words and pseudowords are used. In addition, the increased sensitivity of lexical decision to repeated pseudowords when speeded decisions are required was predicted by the relative familiarity hypothesis. Furthermore, the use of semantic information as a source of additional constraint in lexical decision was an initial assumption of the relative familiarity hypothesis, and hence the observed facilitation of lexical decisions by meaning is consistent with this prediction. Finally the opposite effects of case information found for lexical decisions with words and pseudowords is easily interpreted by the relative familiarity hypothesis, and without added assumptions, as a decision bias based upon the subjects greater experience with some typographies.

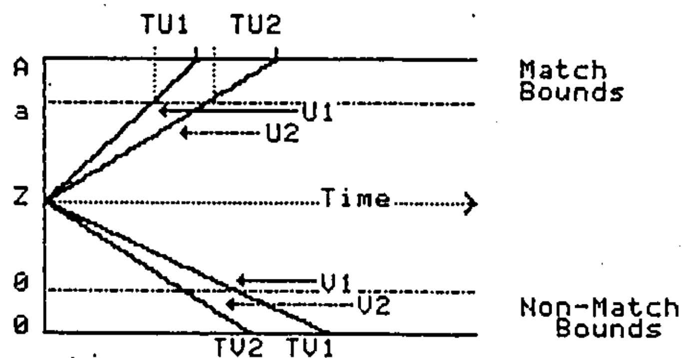
A Common Process for Latency and Accuracy

In the discussion following Exp. 5 it was proposed that response latency and response accuracy could be usefully viewed in terms of normalized distributions and that both of these response measures were based upon the same set of processes. The following discussion presents an argument in support of this claim.

Time dependent processing in the Relative Familiarity Judgment model is modeled by the Random Walk Diffusion Process. In the simplest form the random walk predicts that the effect of word frequency and the manipulation of response accuracy, such as those used in Exp. 2, should result in an interaction for latency. This prediction is similar in form to the interaction described by Becker and Killion (1977). This similarity may be seen in Figure 14. In this figure, the relative latencies for frequent and rare words are determined by the intercept of the diffusion slopes, for frequent and rare words with the decision barrier. In the [illegible] walk model, information about similar and discrepant features for frequent words is assumed to be accumulated at a greater rate than for rare words. Hence the slope representing this growth of item information in

Fig. 14

EXPECTED LATENCIES in the simplest version of the RANDOM WALK PROCESS.



Where: A > a > Z > 0 : and
U and V are slopes based on similarity.

: U1 = slope for FREQUENT WORDS
: U2 = slope for RARE WORDS

and : 'A' is the decision barrier for a CAUTIOUS DECISION

and : 'a' is the decision barrier for a SPEEDED DECISION.

The average decision latency increases from left to right, and is determined by the intercepts of the slopes with the decision barriers. The vertical dotted lines from the SPEEDED decision barrier are provided as a convenience for comparison of the two decison criteria. Note that the estimated decision time for RARE WORDS (TU2) increases more than the estimated decision time for FREQUENT WORDS (TU1).

Figure 14 is steeper for frequent words. As decisions to frequent and rare words require the same ratio of similar to dissimilar features, then the effect of different accuracy instructions, such as that used in Exp. 2, will be to increase or decrease the amount of information required for a positive word identification equally for rare and frequent words by raising or lowering the positive decision barrier. As the rate of approach is different for rare and frequent words, the change from liberal criterion to a consevative criterion should increase decision latency for rare words more than for frequent words, and hence the random walk model predicts an interaction between word frequency and visual information. (While the labels differ, the geometry of the argument is identical to that used in Figure 12.). Hence, in its simplest form the random walk diffusion process makes the same preditions as those described by Becker and Killion (1977) for criterion bias models.

The random walk model predicts additive effects of word frequency and accuracy on decision latency if several standard psychophysical assumptions are made. First, the encoding of relevant information in a lexical decision has a variable and normal distribution of arrival times; and second, the effective value of encoded information is inversely related to the cumulative sum. That is, similarity is some non-linear function of feature matches, such as that provided by the Weber-Fechner function.

In the typical psychophysics experiment, such as that described by Link (1978), it is often reasonable to assume that the rate of comparison, and hence the rate of arrival of relevant information will be some linear function of time. In the Lexical Decision Task, however, it may be inapproriate to assume such a sampling distribution. More appropriate would be a normal distribution of arrival times for relevant information. Hence, the accumuluation of similarity and difference information in the comparison process would also be distributed normally in time. Moreover, this distribution of arrival times would be the same for frequent and rare words.

Frequent and rare words are assumed to differ in the number of features that must be sampled before a set ratio of feature matches to feature non-matches have been accumulated. That is, as frequent words have a higher probability of detecting a feature match given an appropriate feature, then fewer features must be sampled to obtain the information required for a decision. As a result, the average number of primitive features required to accept a frequent word is less than that required to accept a rare word. Notice, however, that it is not that frequent words are responded to on the basis of fewer feature matches. Frequent and rare words require the same amount of information about feature matches but differ in the reliability with which a potential feature match is detected.

In many respects these assumptions are similar to the assumptions underlying Morton's Logogen System. The difference, however, lies in the conceptualization of the bias. For Morton, the bias is located in a specific and specialized word detector and is apparent only in the firing of the logogen, whereas in the current model the bias lies in the increased compatibility of the stimulus information with memories of, or memories for, a particular grapheme and in effect produces an increased rate of feature comparison for well known items. Hence where for Morton, word frequency is entirely a product of <u>Beta</u> in the terms of signal dectecton theory, in the present model word frequency includes differences in <u>d'</u>, as well as in <u>Beta</u>.

An additional assumption for the Random Walk process concerns the scaling of similarity information. It is proposed that information in the comparison process follows a Weber-Fechner law of discrimination, rather than a simple additive function. When similarity is scaled in terms of a Weber-Fechner function, as is usual in psychophysics (i.e., as a logarithmic function) then, within the comparison process, a small change in similarity at low levels of stimulus information will be equivalent to larger changes of similarity at higher levels of stimulus information. The effect of this assumption is to produce a constant and additive bias for frequent words over large changes in stimulus information.

When these assumptions about the temporal distribution of information and the ratio scaling of the similarity information are included in a computer simulation, the time dependent accumulation of information in the random walk process is found to be a non-linear function of information density. The curves plotted in Figure 15 show a computer simulation of the time dependent accumulation of information for frequent and rare words when (a) the availability of feature information is logistically distributed with equal mean and variance for frequent and rare words; and (b) the approach to the decision boundaries is a logarithmic function of the arithmetic difference between feature matches and non-matches. The curves plotted for frequent and rare words differ only in the sampling parameter (probability of a feature match or non-match) used to calculate the above difference score. As can be seen, the slopes associated with frequent and rare words are parallel, non-linear functions; and a change in accuracy mediated by a higher or lower decision boundary, such as was found in Exp. 2, would lead to similar changes in mean response latency for frequent and rare words. Hence, a change in task accuracy should have an additive effect on response times for frequent and rare words.

The curves plotted in Figure 16 show a computer simulation of the

Fig. 15 Time Dependent Processing assuming a Non-Linear Relationship between TIME and STIMULUS INFORMATION.

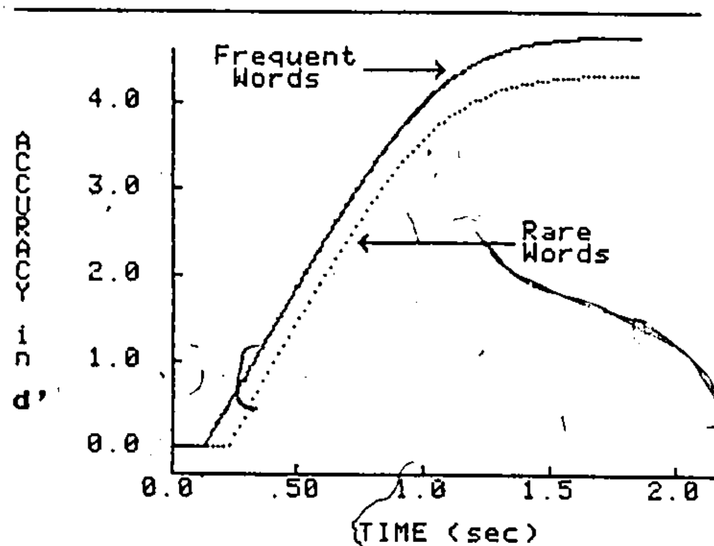
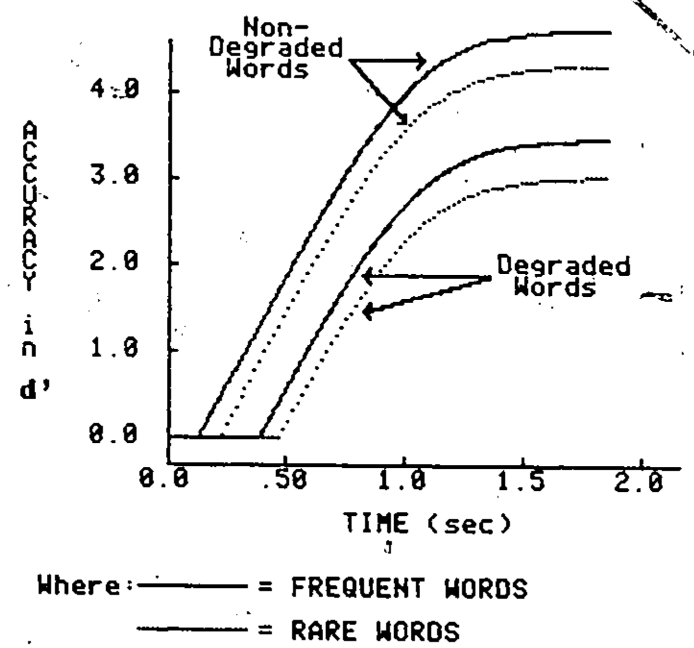


Fig. 16 Time Dependent Processing assuming a Non-Linear Relationship between TIME and the QUALITY of stimulus information.



time dependent accumulation of information for frequent and rare words when, the stimulus is degraded and when it is non-degraded, using the two assumptions discussed previously. It was assumed that degrading a stimulus reduced the availability of feature information equally for frequent and rare words. All other parameters were the same for the degraded and non-degraded stimuli. As can be seen, the expected change in latency introduced by degradation is similar for frequent and rare words; and hence stimulus quality and word frequency should have additive effects on response latency.

The modifications suggested for the random walk process can also be applied to other criterion bias theories. Hence, the Logogen System can be easily modified to account for the observed similarity between mean response times and the mean normalized accuracy of responses. In addition, the time and accuracy curves discussed by McClelland (1979) for analyzing processes in cascade make similar predictions for mean latency and normalized accuracy. In conclusion, from the foregoing discussion of time dependent processing in criterion bias models, it would appear that the additivity in response latency Becker and Killion (1977) argued was only compatible with the Verification Model is expainable within the framework of a criterion bias models. Moreover, criterion bias models can describe additivity in normalized decision accuracy, while the Verification Model cannot.

Before continuing it will be useful to discuss the pseudoword data in Exp. 2 in light of the above distinction. Decisions for repeated pseudowords were delayed, relative to novel pseudowords, by the same amount for both the speed and accuracy instructions, while the relative number of errors for repeated pseudowords was larger with speed instructions. However, as for words, there is an additive effect of instructions on both accuracy and latency when accuracy is considered in terms of normalized accuracy scores (see Figures 10a and 10b). This similarity between latency and normalized accuracy in both word and pseudoword responses reinforces the necessity of examining latency and accuracy as related response measures.

<u>An Independence between Encoding and Representation</u>

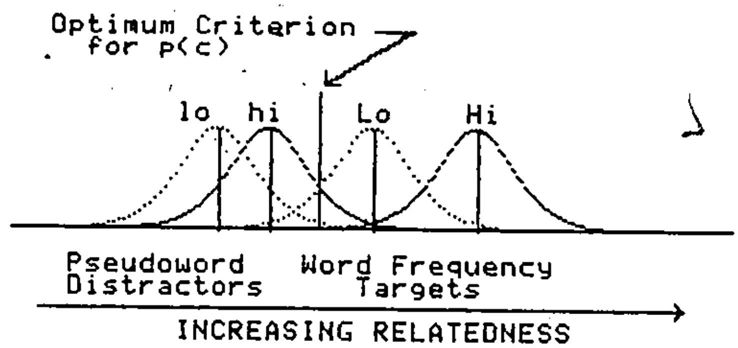
In the previous section latency and accuracy were attributed to a common decision process based on the relatedness dimension and the Random Walk Diffusion Process. Moreover, the random walk process and the relatedness dimension were described as having separable effects on decisions for frequent and rare words. In the following section, the implications of this independence will be developed.

First, changes within the random walk process should have no effect upon relationships maintained by the relatedness dimension. Thus, the changes in processing produced by accuracy instructions, such as the speed and accuracy instructions in Exp. 2, which affect only the parameters of the random walk process, should not affect the average processing difference observed between frequent and rare words. This independence assumes that the relatedness dimension and decision criterion are likely to be similar for the tasks which maintain the same set of targets and distractors.

In order to produce dissimilar frequency effects it would be necessary to increase or decrease the relatedness of the distractors while keeping the word set constant. This change should result in a different choice of the decision criterion, which in turn would change the relative distance between the the criterion and the distributions for frequent and rare words. Changing the distractors should result in a smaller Word Frequency Effect if less related distractors are used, and a larger Word Frequency Effect if more related distractors are used (see Figures 17a and 17b). Such a manipulation was used by James (1975) and Balota and Neely (1980) in a Lexical Decision Task using two groups of subjects. Subjects in both groups were given the same set of frequent and rare words as targets. However, subjects in one group received pronounceable nonwords as distractors, while subjects in the other group received unpronounceable nonwords. For the subjects (in Balota and Neely 1980) who received pronounceable nonwords, the average response latency for frequent and rare words was 632 and 780 msec, respectively, for a difference of 148 msec; while the subjects who received the non-pronounceable nonwords had response times of 598 and 668 msec, respectively, for a difference of 70 msec. Hence, a set of frequent and rare words do not always lead to similar effects of word frequency. A change in the task demands, in terms of target and distractor distinctiveness, can produce non-additive effects of word frequency. Further, as would be expected there were more errors to both words and nonwords with the pronounceable nonword distractors. Thus, the assumption that an easier discrimination will produce a smaller effect of frequency is supported by the data reported by Balota and Neely (1980). Remember though, that an easy task is defined by the nature of the decision space and not, as was seen in Exp. 2, by whether the subjects must make a quick decision. Similarly, changing the visual quality of the display does not change the ease of lexical decisions if the same set of stimuli are used. While letter recognition may be more difficult, and hence lexical decisions delayed, the interpretation of the letter strings still occurs within the same decision space.

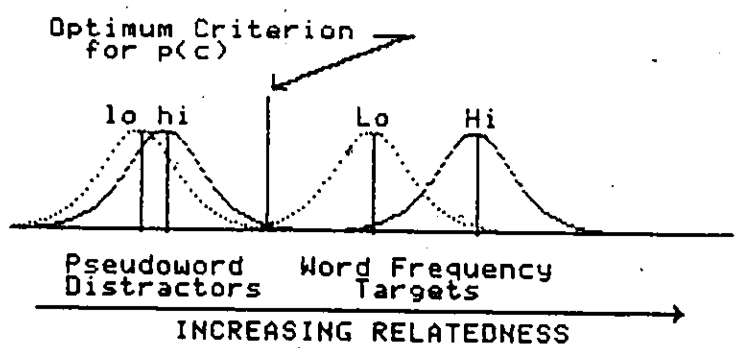
In summary, the relationships described by the relatedness distribution are assumed to be determined jointly by the type of stimuli encountered and the differential knowledge about these stimuli in memory. While

Fig. 17a
High and Low Frequency Words: with 50% Similar Distractors.



The relative RATIO of the criterion to Hi and Lo Frequency Words is LARGE.

Fig. 17b
High and Low Frequency Words: with 25% Similar Distractors.



The relative RATIO of the criterion to Hi and Lo Frequency Words is SMALL.

the nature of the relatedness dimension may vary between tasks, it, like the decision criterion (A) in the random walk process, is a constant within a single experiment.

As an example of this last point, consider a comparison of the Word Frequency Effect observed in the unbiased group in Exp. 1 with that observed in Exp. 5. Both experiments used the same set of words and pseudowords, yet the average size of the Word Frequency Effect in Exp. 1, 48 msec, was almost half that seen in Exp. 5, 88 msec. It would appear, then, that the Word Frequency Effect is not a constant size for a given set of stimuli. The actual size may be relative to both stimulus and task conditions. A possible cause of the different effects of word frequency in the two experiments is that half the items in Exp. 5 were degraded. Yet despite the differences between experiments both degraded and non-degraded words showed the same word frequency difference in Exp. 5. Thus, it was not a question of whether specific stimuli were degraded or not which led to the different effects of word frequency, but apparently the general characteristics of the test environment. While the average response time for the unbiased group was slightly faster than the average response time to non-degraded words in Exp. 5 (561 and 572 msec, respectively), we know from Exp. 2, with speed and accuracy instructions, that a speed-accuracy trade-off is not a sufficient basis for a change in the Word Frequency Effect. Interestingly, the error rate in Exp. 1 (for both words and pseudowords) was larger than the error rate to non-degraded items in Exp. 5. This suggests that the characteristics of the memory set, or the relatedness dimension, used for decisions in the two experiments were different. That is, the memory standard used in the comparison process was different for the two experiments. This makes intuitive sense, of course, as at least half the items in Exp. 5 were visually degraded. Consequently, it seems reasonable to suppose that the stimuli would require a more extensive analysis because of this degradation. It would appear, however, that subjects once committed to extensive analysis had to apply the same standard for degraded and non-degraded words. In short, while the relatedness dimension may vary across experiments (but not necessarily, given the results of Exp. 2) it should be a constant within any given experiment.

In summary, while the Word Frequency Effect is often additive across experimental manipulations (e.g., Exp.s 2 and 5), it is possible, nonetheless, to find non-additive effects. This non-additivity may be accomplished by a) holding the relatedness relations for words constant while manipulating the level of distractor relatedness; and b) changing the average availability of stimulus information within an experiment, such as by a manipulation of the average visual quality. Another method for changing the relatedness dimension is semantic context, and will be discussed in the next section.

Semantic Information and Lexical Decisions

To this point a relative familiarity judgment has been discussed mainly in terms of visual familiarity. This however has been a simplification for the purposes of clarity. Relative judgments are not an isolated property of visual-grapheme processing; they are one aspect of an incomplete similiarity comparison, and hence are likely present in most non-reflective decisions. As seen, semantic information as well as visual familiarity can determine lexical decisions, and in Relative Familiarity Judgment model this is explained by assuming that semantic information is included in a relative decision. Specifically, it is suggested that as semantic constraints are used in representing stimuli, they also contribute to the relatedness dimension; and hence semantic information is seen as a direct source of information in a relative judgment.

An aspect of semantic information not discussed to this point is the effect of semantic context on lexical decisions as introduced by Meyer, Schvaneveldt and Ruddy (1975). Semantic context in this paradigm entails the presentation of a word related to the target contiguous to a target word. Typically, lexical decisions for the target word are facilitated by this contiguity (Schvaneveldt, Meyer & Becker, 1976; Fischler, 1977; Becker & Killion, 1977). For example, lexical decisions about NURSE would be faster and more accurate when preceded by DOCTOR than by LAWYER. At least initially, this would seem to provide strong support for the present argument of multiple constraints in the Lexical Decision Task. This is true, however, only if semantic context is processed during the acutal recognition of a target word. If, as is suggested by Morton (1969) and Becker and Killion (1977), semantic context speeds lexical decision by pre-biasing the visual representation of the target word, then semantic processing does not need to be considered as active during lexical decision. That is, the active word decision may be considered as occurring exclusively at the level of visual processing when the effect of context is to pre-biases this processor. Because of this potential difficulty in determining the manner in which semantic context contributes to processing, the effects of semantic context in this form have been ignored to present. The following section presents a potential method for selecting between an explanation of semantic context in terms of a concurrent processing of semantic information or a semantic pre-biasing of the visual processing.

Semantic context in the present perspective can facilitate lexical decisions in two fashions. First, it can facilitate decisions at a semantic level by emphasizing certain semantic features in the decision standard; and secondly it can influence the expectation of a particular word's visual features.

A temporary change in semantic information should affect processing in much the same way as other changes in memory information. That is, it should increase the probability of a feature match in the comparison process, and hence it should increase the speed and reliability of a positive decision. Moreover, it should operate on performance in a fashion similar to the changes in memory information discussed earlier, such as frequency and repetition. As with the other memory changes semantic context will have a larger effect in facilitating similar targets than in inhibiting dissimilar targets. The justification for this claim is that in a large discrimination set, such as the relatedness dimension, the number of features shared in common by the memory information and a given target is far smaller than the number of features not shared in common. Hence incrementing the number of similar and dissimilar features by a similar amount will primarily affect positive feature matches. As decisions depend upon the ratio of feature matches to feature nonmatches then adding information generally increases discriminability. Hence, similar to frequency and repetition information, semantic context is primarily facilitory in operation.

The second possible mode of action for semantic context resembles the spreading activation explanation of context effects (see Tweedy, Lapinski & Schvaneveldt, 1977). Not only are semantic constraints changed, the nature of the visual representation is also affected. However, as the previous discussion this change in memory representation mainly facilitates the primed target, or stimuli similar to it, and should leave decisions for dissimilar targest largely unaffected.

While the benefits of semantic context, word frequency and repetition are attributed to a similar change in memory representation, there are important differences between the three conditions, the primary distinction being the potential for strategic manipulation of these conditions. Word frequency is a diffuse and long term aspect of word experience and because of this is not open to manipulation by expectation. That is, it seems unlikely that a subject could intentionally emphasize frequent or rare words by manipulating the relatedness dimension. Semantic context and repetition, on the other hand, are easily subjected to such manipulation. Tweedy et al. (1977) showed that the facilitation from semantic context increases as the proportion of related pairs encountered during a Lexical Decision Task increases. This change in facilitation could be due to either of the two processes discussed above but in either case it reflects a change in the extent to which information is used in lexical decision.

While a similar manipulation of strategic information has not been shown for repetition, there is no reason that it should not show a similar flexibility. An interesting test of this prediction would be to apply the Episodic Priming technique used by McKoon and Ratcliff (1979) with the manipulation of the proportion of primed pairs used by Tweedy et al. (1977). McKoon and Ratcliff. (1979) found that newly learned paired associates (e.g., MARBLE-HOUSE) yielded as large a facilitation in lexical decisions as did semantic associates (e.g., BABY-CHILD) when tested in sequence. If increasing the frequency of learned paired associates in a Lexical Decision Task produced the same pattern of increasing facilitation reported by Tweedy et al. (1977) ~~for~~ semantic associates, then strategic priming would be shown for repetition.

Evidence of strategic priming has been found in the Word Identification Task. Jacoby (1982a) found that the benefit of a prior presentation in word identification depended upon the type of words in the test list. The relative benefit of a prior presentation increased when the test list contained more old words than new words, whereas the relative benefit of a prior presentation decreased when the test list contained more new then old words. It is likely, then, that information gained from a prior presentation can lead to effects of strategic priming similar to those shown by semantic context.

The present claim is that semantic context can show benefits of strategic priming while word frequency will not. This has an important implication for any explanation of the Lexical Decision Task. An often quoted difference between the two effects of word frequency and semantic context is that semantic context interacts with visual quality (Meyer, Schvaneveldt & Ruddy, 1974; Becker & Killion, 1977) while word frequency does not (Stanners, Jastremvske & Westbrook, 1975; Becker and Killion, 1977). A possible explanation for this comes from the preceding argument of strategic effects. Within the relative familiarity hypothesis, decisions based upon a similar relatedness continuum will result in similar effects of memory information, when performance is analyzed in normalized accuracy distributions. This represents the case when word frequency and visual quality are tested, and when word frequency is tested across different levels of accuracy instruction. Changing the relatedness dimension, however, can lead to different effects of identical word frequency information, e.g., Balota and Neely (1980). Consider now semantic context. In a manipulation of visual quality, the semantic facilitation increases as the stimulus is degraded. It might be expected that as the the visual quality decreases, the strategic emphasis on semantic context will also increase. In this case, the relatedness continuum should also change, and hence the amount of facilitation shown for semantic context should also change. As you will remember, changing the relatedness continuum used in a lexical decision also changes the size of the Word Frequency Effect. In short, it is because semantic context lends itself to strategic effects and word frequency does not, that the two factors lead to different results when visual quality is degraded.

The present description leads to testable predictions. Specifically, if the relatedness continuum is biased by semantic context then there should be some evidence of this bias. The proper test for this memory bias would entail examining decisions for alternate targets which share features with the changed information. Remember, it is assumed that a change in memory information will in general result in increases for both <u>d'</u> and <u>Beta</u>. Thus, the overall discriminability of a primed target relative to a similar distractor should increase, at the same time the relative proportion of errors made to a similar distractor should increase.

While the above prediction has not been tested directly, a series of experiments reported by Schvaneveldt and McDonald (1981) offer a close approximation. In this test semantic context was examinded in two conditions, a conventional Lexical Decision Task with high accuracy and a Lexical Decision Task where the stimuli were flashed tachistoscopically and followed by a mask. In the latter test, only accuracy was gathered. However, as the overall error rate in this procedure was similar to that seen in Exp. 2 for the speed condition, the task should be a fair assessment of lexical decisions about a degraded stimulus. The critical factor in Schvaneveldt and McDonald (1981) was the use of similar distractors in both the primed and unprimed conditions, as well as the set of primed and unprimed words. There were four relevant comparisons, a primed word (e.g., LION-TIGER), a primed distractor (e.g., LION-TIGAR), an unprimed word (e.g., RAIN-TIGER), and an unprimed distractor (e.g., RAIN-TIGAR). The subjects were to respond nonword for TIGAR, and word for the correct spelling. Figures 18a and 18b show the normalized accuracy scores for each condition, for the easy and difficult Lexical Decision Tasks. First, note that the difference between the primed word and unprimed word conditions increased for a degraded presentation. This is opposite to what is typically found for word frequency when stimuli are degraded (Stanners et al. 1975, Exp. 5) and when a speeded judgment is required (Exp. 2). That is, the normalized accuracy scores show an interaction for semantic context but a main effect for word frequency. This of course is the same pattern of results typically found with decision latencies and on the basis of the previous discussion relating accuracy and latency it is suggested that the parallel effects of degradation on accuracy and latency is based upon a shared and common process.

The second point of interest in Figures 18 is the rejection of misspelled words. Distractors similar to words yielded relatively more errors when preceded by the similar word's prime. Moreover, as for the words, the manipulation of context had the largest effect when the stimuli were degraded. This increase in normalized errors is opposite to the additivity of normalized errors seen for pseudowords in Exp. 2, for the speed and accuracy instructions. Within the present model, the relative change in accuracy for primed words and similar nonwords implies a change in the characteristics of the relatedness dimension when primes are used with visually degraded stimuli. This restructuring of memory is a feasible strategy for semantic information but not

Fig. 18a Semantic Facilitation with NO TIME CONSTRAINTS. (from Schvaneveldt & McDonald, 1981)

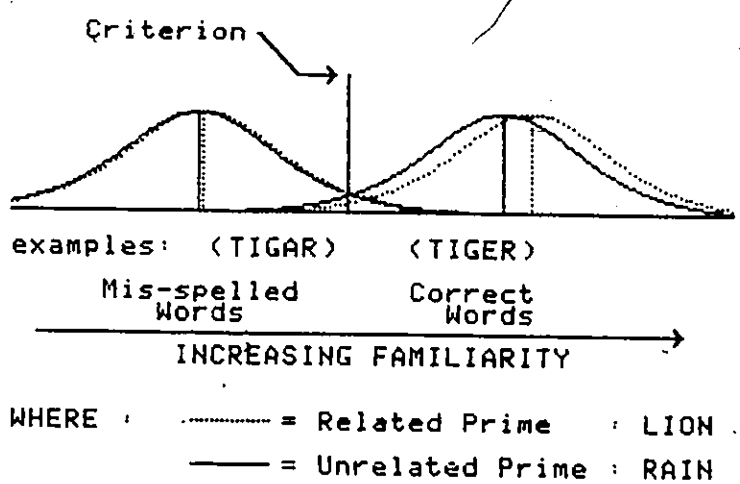
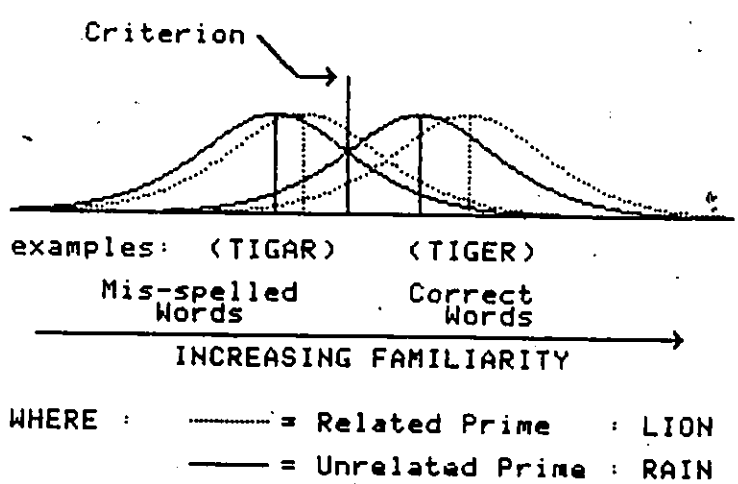


Fig. 18b Semantic Facilitation with TACHISTOSCOPIC PRESENTATION. (from Schvaneveldt & McDonald, 1981)



for word frequency.

There are three important points to be made from the preceding discussion. First, the effects of strategic priming probably account for the interaction of semantic context with visual quality. Second, as discussed in the introduction, the effect of a memory bias in a complex test environment produces a change in both the estimate of <u>d'</u> and <u>Beta</u>. This point cannot be emphasized too strongly, all too often memory changes are considered to be either a <u>d'</u> shift or a <u>Beta</u> shift. In point of fact a change in memory is neither. It is the test environment and the type of information tested that determines how the change in information can be used, and whether a change in memory will function primarily as a <u>d'</u> of a <u>Beta</u> change. The third point is related to the second point, the effective change in memory information, and hence the bias due to semantic context, was different in the two test environments shown in Figures 18a and 18b.

Earlier it was claimed that semantic context may operate in two fashions, at the level of semantic features and at the level of visual features. Both of these strategies have their respective advantages and disadvantages. Priming at the level of semantic features will generalize across semantically related but visually distinct words. Priming at this level, though, is not especially useful with reduced visual information. Priming at the level of visual features, on the other hand, should facilitate processing of a degraded stimulus with a cost of greater sensitivity to visually similar distractors. If priming were solely for visual features then there should be no generality of semantic context. In short, it is proposed that the change in strategic priming observed by Schvaneveldt and McDonald (1977) reflected a change in the type of information tested in addition to a change in the emphasis of semantic context.

There are several experiments which might test this prediction, but unfortunately there is no particular reason that priming should occur exclusively at one level or the other. For this reason the difference between conditions may be one of emphasis rather than of presence or absence of information at either level. As a result, without an explicit formula for generating these weightings, the relative familiarity hypothesis can only serve as a rough guide for making predictions.

In conclusion, it is suggested that the Relative Familiarity Judgment model developed earlier can also describe the function of semantic context in the Lexical Decision Task. Furthermore, it is proposed that semantic context like word frequency is mediated by difference in word knowledge, and that both these memory based constraints on judgment share common encoding and decision processes. The two differ, however, in terms of the type of information emphasized and the ease of their strategic manipulation. Finally, it is proposed that the last feature, the ease of strategic manipulation, accounts for the major experimental difference found for word frequency and semantic context. Thus, where other theories locate word frequency and semantic context at separate stages, or in different processes, the present account stresses the similarity of representation and process, and emphasizes the factors that manipulate memory based information.

Conclusions.

In the introduction a contrast was drawn between lexical store models of the Lexical Decison Task and the relative familiarity model. The lexical store models make several assumptions about the nature of the memory investigated by the Lexical Decision Task, such as that only meaningful words are represented in lexical memory, and that the information used for lexical access is independent of variations in visual form. On the basis of the present results it is clear that both these assumptions are incorrect. Pseudoword strings previously seen during a visual letter search task are likely to be falsely recognized as real words (Exp. 2 and 3). Moreover, in Exp. 3 it was found that lexical decisions are systematically biased in terms of the visual form of the graphemic strings. Decision for words were facilitated when test stimuli were in a lower case typography, while decisions for pseudowords were hurt by a lower case presentation. From this pattern of results, it is clear that some aspects of the "physical" or visual form of the grapheme match the internal word knowledge better than other forms. Hence, it can be concluded that lexical decisions are not a pure measure of visual, graphemic knowledge. Lexical decisions reflect in varying degrees knowledge about all aspects of word experience.

Another assumption, that the speed or accuracy of lexical access in independent of task demands is clearly not in accord with the results of Exp. 2 or of Balota and Neely (1980). Speed instructions in Exp. 2 could be compared to word recognition while skimming, while accuracy instructions could be compared to reading for detail. Lexical decisions do not necessarily measure access time, rather it would seem they measure the time to acheive a task specified level of confidence. This sensitivity of the speed of lexical decisions to task demands has important implications. In a recent paper Glanzer and Ehrenreich (1979) report that lexical decisions to frequent words may depend upon the stucture of the "internal lexicon". In a series of experiments they found that the relative advantage in decision latency for frequent words over rare words increased when frequent and rare words were presented in blocked lists, and decreased when they were presented in mixed lists. They interpret the advantage of blocking word frequency to two "internal lexicons", a complete lexical memory used for mixed trials, and an abridged (frequent words, only) lexical memory used for frequent words in the blocked trials. On the basis of Exp. 2, two "internal lexicons" is clearly inadequate in accounting for speeded lexical decisions. It seems more likely that subjects shifted their response criterion between blocked and mixed trials more for frequent words than for rare words. On the basis of Exp. 2 such a strategy would have a large benefit for decision latency with frequent words and only a small cost in decision accuracy. In conclusion, the speed of lexical access is not invariant, rather it depends upon task and stimulus demands.

Further, the relative increase in false alarms to repeated pseudowords when decisions are speeded suggests that partial information is available in lexical decisions. That is, while detection of a word may appear to be all or none, it is the decision process that produces this characteristic. Access to memory would appear to be graded and decisions all or none.

There are several implications of the above conclusions. First, the initial stage in any use of a word is to make contact with specialized memory information, not a specialized lexical memory. Second, the Lexical Decision Task is an indirect measure of this initial stage. Lexical decisions appear to be familiarity judgments about word stimuli out of context, while most tasks using words stress meaningful interpretations within a context. Hence, the Lexical Decision Task is not a direct and unbiased measure of the word processing skills used in reading. Undoubtedly there are similarities in the origin of skills used in reading and in lexical decisions, but the type of stimuli used as targets and distractors, the nature of task demands, and variations in the visual form of the stimuli may all lead to significant differences in the Lexical Decision Task that are irrelevant to normal reading.

References

Aderman, D., & Smith, E. E. Expectancy as a determinant of functional units in perceptual recognition. Cognitive Psychology, 1971, 2, 117-129.

Atkinson, R. C., & Joula, J. F. Factors influencing speed and accuracy of word recognition. In S. Kornblum (Ed.) Attention and performance IV. New York: Academic Press, 1973.

Balota, J., & Neely, J. H. Levels-of-processing and word-frequency effects in episodic and semantic memory. Private communication from Dr. J. H. Neely, March, 1980.

Becker, C. A. Allocation of attention during visual word recognition. Journal of Experimental Psychology: Human Perception and Performance, 1976, 2, 556-566.

Becker, C. A. Semantic context and word frequency effects in visual word recognition. Journal of Experimental Psychology: Human Perception and Performance, 1979, 5, 252-259.

Becker, C. A., & Killion, T. H. Interaction of visual and cognitive effects in word recognition. Journal of Experimental Psychology: Human Perception and Performance, 1977, 3, 389-401.

Broadbent, D. E. Word-frequency effect and response bias. Psychological Review, 1967, 74, 1-15.

Broadbent, D. E. Decision and stress. London: Academic Press, 1971.

Broadbent, D. E. The hidden preattentive process. American Psychologist, 1977, 32, 109-118.

Broadbent, D. E., & Broadbent, M. H. P. Some further data concerning the word frequency effect. Journal of Experimental Psychology: General, 1975, 104, 297-308.

Brooks, L. R. Non-analytic concept formation and memory for instances. In Rosch, E. and Lloyd, B. (Eds.), Cognition and Categorization. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1978.

Brooks, L. R. On the insufficiency of aanalysis. Manuscript submitted for publication, 1982.

Carr, T. H., Davidson, B. J., & Hawkins, H. L. Perceptual flexibility in word recognition: Strategies affect orthographic computation but not word access. Journal of Experimental Psychology: Human Perception and Performance, 1978, 13, 153-156.

Catlin, J. In defense of the sophisticated-guessing theory. Psychological Review, 1973, 80, 412-416.

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, 1975, 104, 268-294.

Eriksen, C. W., Pollack, M. D., & Montage, W. E. Implicit speech: Mechanism in perceptual encoding? Journal of Experimental Psychology, 1970, 84, 502-507.

Fischler, I. Semantic facilitation without association in a lexical decision task. Memory & Cognition, 1977, 5, 335-339.

Forbach, G. B., Stanners, R. F., & Hochhaus, L. Repetition and practice effects in a lexical decision task. Memory & Cognition, 1974, 2, 337-339.

Frederiksen, J. R. Assessment of perceptual, decoding and lexical skills and thier relation to reading proficiency. In A. M. Lesgold, J. W. Pellegrino, S. D. Fokkema, & R. Gleaser (Eds.) Cognitive Psychology and Instruction. New York: Plenum Press, 1978.

Frederikson, J. R., & Kroll, J. F. Spelling and sound: Approaches to the internal lexicon. Journal of Experimental Psychology: Human Perception and Performance, 1976, 2, 361-379.

Garner, W. R. The role of configuration in the identification of visually degraded words, Memory & Cognition, 1981, 9, 445-452.

Glanzer, M., & Bowles, N. Analysis of the word-frequency effect in recognition memory. Journal of Experimental Psychology: Human Learning and Memory, 1976, 2, 21-31.

Glanzer, M., & Ehrenreich, S. L. Structure and search of the internal lexicon. Journal of Verbal Learning and Verbal Behavior, 1979, 18, 281-398.

Green, D. M., & Swets, J. A. Signal detection theory and psychophysics. New York: Wiley, 1966.

Jacoby, L. L. Perceptual enhancement: Persistent effects of an experience. Accepted for publication in the Journal of Experimental Psychology: Learning, Memory, and Cognition, 1982. (a)

Jacoby, L. L. Remembering the Data: Analyzing interactive processes in reading. Accepted for publication in the Journal of Verbal Learning and Verbal Behavior, 1982. (b)

Jacoby, L. L., & Craik, F. I. M., Effects of elaboration of processing at encoding and retrieval: Trace distinctiveness and recovery of initial context. In L. S. Cermak & F. I. M. Craik (Eds.), Levels of processing and human memory. Hillsdale, N.J.: Earlbaum, 1979.

Jacoby, L. L. & Dallas, M. On the relationship between autobiographical memory and perceptual learning. Journal of Experimental Psychology: General, 1981, 110, 306-340.

Jacoby, L. L., & Witherspoon, D. Remembering without awareness. Canadian Journal of Psychology, 1982, 36, 300-324.

James, C. T. The role of semantic information in lexical decisions. Journal of Experimental Psychology: Human Perception and Performance, 1975, 104, 130-136.

Kirsner, K., & Smith, M. C. Modality effects in word identification. Memory & Cognition, 1974, 2, 637-640.

Kleiman, G. M. Speech recoding in reading. Journal of Verbal Learning and Verbal Behavior, 1975, 14, 323-339.

Kolers, P. A., & Smythe, W. E. Images, symbols, and skills. Canadian Journal of Psychology, 1979, 33, 158-184.

Link, S. W. The relative judgment theory of two choice response time. Journal of Mathematical Psychology, 1975, 12, 114-135.

Link, S. W. The relative judgment theory of the psychometric function. In J. Requin (Ed.) Attention and Performance VII. New York: Lawrence Earlbaum Associates, 1978.

Link, S. W. Private communication from Dr. S. W. Link, May, 1981.

Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. Functions of graphemic and phonemic codes in visual word recognition. Memory & Cognition, 1974, 2, 309-321.

Meyer, E. D., Schvaneveldt, R. W., & Ruddy, M. G. Loci of contextual effects on visual word recognition. In P. M. A. Rabbit & S. Dornic (Eds.) Attention and Performance V. New York: Academic Press, 1975.

McClelland, J. L. Preliminary letter identification in the perception of words and nonwords. Journal of Experimental Psychology: Human Perception and Performance, 1976, 2, 80-91.

McClelland, J. L. On the time relations of mental processes: A framework for analyzing processes in cascade. Psychological Review, 1979, 86, 287-330.

McClelland, J. L., & Rumelhart, D. L. An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. Psychological Review, 1981, 88, 375-407.

McKoon, G., & Ratcliff, R. Priming in episodic and semantic memory. Journal of Verbal Learning and Verbal Behavior, 1979, 18, 463-480.

Morton, J. Interaction of information in word recognition. Psychological Review, 1969, 76, 165-178.

Morton, J. A functional model for memory. In D. A. Norman (Ed.) Models of human memory. New York: Academic Press, 1970.

Morton, J. Facilitation in word recognition: Experiments causing change in the logogen model. In P. A. Kolers, M. Wrolstrad & H. Bouma (Eds.), Processing of Visible Language, Vol. 1. New York: Plenum, 1977.

Morton, J., & Broadbent, D. E. Passive vs. active recognition models or Is your homunculus really necessary? In W. Wathen-Dunn (Eds.), Models for the perception of speech and visual form. Cambridge, Mass.: M.I.T. Press, 1967.

Morton, J., & Paterson, K. A new attempt at an interpretation, or, an attempt at a new interpretation. ~~In M.~~ Coltheart, K. Patterson, & J. C. Marshall (Eds.), Deep Dyslexia. London: Routledge & Kegan Paul, 1980.

O'Connor, R. E., & Forster, K. I. Criterion bias and search sequence bias in word recognition. Memory & Cognition, 1981, 9, 73-92.

Pavio, A. & O'Neil, B. J. Visual recognition thresholds and dimensions of word meaning. Perception & Psychophysics, 1970, 8, 273-275.

Posner, M. I., Boise, S. J., Eichelman, W. H., & Tailor, R. L. Retention of visual and name codes of single letters. Journal of Experimental Psychology Monograph, 1969, 79(1, Pt. 2).

Posner, M. I., Klein, R., Summers, J., & Buggie, S. On the selection of signals. Memory & Cognition, 1973, 1, 2-12.

Ratcliff, R. A theory of memory retrieval. Psychological Review, 1978, 85, 59-108.

Ratcliff, R., & Murdock, B. B., Jr. Retrieval processes in recognition memory. Psychological Review, 1976, 83, 190-214.

Rayner, K., & Posnansky, C. Stages of processing in word identification. Journal of Experimental Psychology: General, 1978, 107, 64-80.

Rubenstein, H., Lewis, S. S., & Rubenstein, M. A. Evidence for phonemic recoding in visual word recognition. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 645-657.

Rumelhart, D. E., & McClelland, J. L. An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. Pschological Review, 1982, 89, 60-94.

Scarborough, D. L., Cortese, C., & Scarborough, H. S., Frequency and repetition effects in lexical memory. Journal of Experimental Psychology: Human Perception and Human Performance, 1977, 3, 1-17.

Scarborough, D. L., Gerard, L., & Cortese, C. Accessing lexical memory: The transfer of word repetition effects across task and modality. Memory & Cognition, 1979, 7, 3-12.

Shulman, H. G., Hornak, R., & Sanders, E. The effects of graphemic and semantic relationships on access to lexical structure. Memory & Cognition, 1978, 6, 115-123.

Schvaneveldt, R. W., & McDonald, J. E. Semantic context and the encoding of words: Evidence for two modes of stimulus analysis. Journal of Experimental Psychology: Human Perception and Performance, 1981, 7, 673-687.

Schvaneveldt, R. W., Meyer, D. E., & Becker, C. A. Lexical ambiguity, semantic context, and visual word recognition. Journal of Experimental Psychology: Human Perception and Performance, 1976, 2, 243-256.

Selfridge, O. G. Pandemonium: A paradigm for learning. In The mechanisation of thought processes. London: H. M. Stationary Office, 1959.

Shephard, R. Recognition memory for words, sentences and pictures. Journal of Verbal Learning and Verbal Behavior, 1967, 6, 156-163.

Stanners, R. F., Jastrzembski, J. E., & Westerbrook, A. Frequency and visual quality in a word-nonword classification task. Journal of Verbal Learning and Verbal Behavior, 1975, 14, 259-264.

Stanners, R. F., Neison, J. J., & Painton, S. Memory representation for prefixed words. Journal of Verbal Learning and Verbal Behavior, 1979, 18, 737-743.

Sternberg, S. Two choice operations in character recognition: Some evidence from reaction-time measurements. Perception & Psychophysics, 1967, 2, 45-53.

Thorndike, E., & Lorge, I. The teacher's word book of 30,000 words. New York: Columbia University, Teacher's College Press, 1944.

Tweedy, J. R., Lapinski, R. H., & Schvaneveldt, R. W. Semantic-context effects on word recognition: Influences of varying the proportion of items presented in an appropriate context. Memory & Cognition, 1977, 5, 84-89.

Vokey, J. R. & Brooks, L. R. Taming the clever unconscious: Analogic and abstractive strategies in artificial grammar learning. Accepted for publication in Cognition, 1982.

APPENDIX A

The following study was an early attempt to experimentally manipulate meaning in a Lexical Decision Task. The Levels of Processing manipulation was examined in both a Lexical Decision Task and in a typical recognition test. The results showed no effect of the levels manipulation in lexical recognition. However, there was only a minimal effect of a prior presentation. As the test stimuli were mainly, frequent words adopted from a Levels of Processing task, it was assumed that the small effect of a repetition which is typical for frequent words minimized the chances of finding an effect due to the semantic manipulation. On the other hand, the levels manipulation had the expected robust effect in the test of recognition memory. On the basis of this finding, it was felt unnecessary to include a test of recognition memory in Experiment 3.

The general procedure used in this pilot study involved three types of Phase I orienting tasks, and then tested for different levels of tranfer in a subsequent Phase II, Lexical Decision Task. A surprise recognition test was given to each subject following the Lexical Decision Task.

Method

Subjects. Twenty four McMaster undergraduates received an hour's course credit for participating in the experiment.

Materials and Procedure. The word stimuli and Phase I orienting questions, used in Experient 2, were taken from Jacoby and Dallas (1981). The words, which were not controlled for frequency, were 5-letter nouns. Three types of questions were asked for each word, each requiring a yes or a no answer; the ordering of type of question and the type of answer were randomized. The three types of question were: (1) letter search questions (e.g., Contains the letter 'B'?); (2) rhyme questions (e.g., Rhymes with glass?); (3) semantic questions (e.g., Is a rodent?). Phase I presented 60 words (20 words in each of three orienting tasks (10 yes and 10 no, per task). The appropriate counterbalancing was performed between subjects, such that each word was seen equally often in each of the six conditions (three tasks times two responses), and as a new word.

The Phase II, Lexical Decision Task contained 160 targets; 80 words and 80 5-letter pseudowords; of the 80 words, 60 were repeated and 20 were experimentally novel words. The procedure used in the Lexical Decision Task was similar to that used in Exp. 1. The word stimuli were counterbalance across subjects, as old and new words.

At the end of the Lexical Decision Task, each subject was given a surprise, episodic recognition test in order to assess the Levels of Processing manipulation. The recognition test consisted of 240 5-letter nouns, 80 of which were repeated words from the Lexical Decision Task. All the words were typed on a single sheet of paper, and the subjects were instructed to circle only the words they remembered as having seen in the Phase I tasks.

Results and Discussion

The results of the Phase II, Lexical Decision Task are presented in Table A, while Table B presents the results for the test of recognition memory.

As can be seen in Table A, there was only a small benefit of a prior presentation in the Lexical Decision Task, although this difference was significant. However, there were no significant differences in either the response latency or accuracy attributable to the type of the Phase I, orienting

TABLE A

Reaction times (in msec) and Percent Accuracy for Words and Pseudowords, in Appendix A.

	PHASE II RESPONSE given:			
	Phase I RESPONSE			
	'YES'		'NO'	
Phase I task:	RT	ACC.	RT	ACC.
LETTER SEARCH	658	(98.0)	666	(98.1)
RHYME JUDGEMENT	653	(99.2)	675	(97.7)
SEMANTIC JUDGEMENT	649	(99.6)	675	(97.5)

	NEW WORDS		PSEUDOWORDS	
	RT	ACC.	RT	ACC.
NEW IN PHASE II:	673	(93.7)	750	(96.6)

task. However, the overall distribution of latency and errors are compatible with predictions expected from the Levels of Processing perspective. It may be that the overall low level of transfer hide any difference due to the type of Phase I, orienting task.

As was expected, there was a large effect of the Phase I, orienting task in the recognition data, see Table B. The false alarm rate for the 160 distractors was 3.4%. As predicted, the semantic judgements, in Phase I, led to higher overall recognition; although there was not a general benefit of rhyme judgements over letter search questions. Also, there was in general higher recognition for yes responses than no responses, except for the letter search questions where the reverse is true. The greater memory retention for yes responses is a typical finding in a Levels of Processing experiment (see Craik & Tulving, 1975), especially for semantic questions.

In summary, the recognition test showed the usual effect of a Levels of Processing manipulation; while the Lexical Decision Task revealed no significant differences for these manipulations. However, given the small benefits shown for repeated words, and the general similarity of reponses in the Lexical Decision Task and the recognition task, it could not be concluded that contextual information, gained from a prior presentation, has no effect within lexical decision.

#150a

TABLE B

Episodic Recognition scores, for Appendix A, given Percent Accuracy.

	PHASE II RESPONSE given: Phase I RESPONSE	
	'YES'	'NO'
Phase I task:	ACC.	ACC.
LETTER SEARCH	48.8	60.0
RHYME JUDGEMENT	58.3	46.3
SEMANTIC JUDGEMENT	76.3	57.9

	NEW WORDS FALSE ALARMS
NEW IN PHASE II:	3.2

CHANCE PERFORMANCE WAS 33.3 %

Significance in Group comparisons in Recognition.

S/Y	P/N	R/Y	S/N	P/Y	R/N	NEW
$p < .001$			$p < .05$			

Where: S = Semantic Judgement in Phase I
R = rhyme Judgement in Phase I
P = Letter Seach in Phase I

and Y = a 'YES' Response in Phase I
N = a 'NO' Response in Phase I.