FREQUENCY JUDGMENTS FOR RELATED AND UNRELATED EVENTS

## FREQUENCY JUDGMENTS FOR RELATED AND UNRELATED EVENTS

By

# GRANT THOMAS HARRIS, 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

June 1981

DOCTOR OF PHILOSOPHY (1981) (Psychology)

### McMASTER UNIVERSITY Hamilton, Ontario

TITLE: Frequency Judgments for Related and Unrelated Events AUTHOR: Grant Thomas Harris B.Sc. (University of Toronto) SUPERVISOR: Professor Ian Begg NUMBER OF PAGES: xi, 136

### Abstract

This thesis examines people's estimates of the number of times events have occurred. Specifically, the thesis investigates how frequency estimates for pairs of related events differ from estimates for pairs of unrelated events. Previous research on "illusory correlation" (Chapman, 1967) has led to the conclusion that people show a bias that causes them to systematically and grossly overestimate the frequency or correlation of related pairings relative to unrelated The introduction of this thesis presents some empirical and pairings. theoretical grounds to question this characterization of "illusory correlation". The contention is that the theories and the existing body of research about frequency estimation are at odds with the conclusion that there is an overall bias. The introduction develops a theoretical view which predicts that frequency estimates for related pairs will show lower sensitivity and worse discriminability than estimates for unrelated pairs. There are also good reasons to suppose that related and unrelated pairs will not differ in the overall average magnitude of the frequency estimates they each receive.

An important consequence of differences in sensitivity or frequency discrimination is that such differences can look like magnitude differences (a bias) if only a small range of actual frequencies is examined. It is possible that the characterization of illusory correlation as a response bias resulted from a failure to examine a wide enough range of actual frequencies. The first experiment demonstrates that the important difference between frequency estimates for related and unrelated events is lower sensitivity for estimates of related pairings. The second experiment provides evidence that this sensitivity difference occurs because subjects treat related and unrelated pairs differently during study. Basically, this difference in encoding strategy is characterized as attention to the general, categorical or semantic features of related events, and attention to spatial, temporal, episodic characteristics of unrelated events. Because general, semantic encoding is less useful as a basis for later frequency judgments than is specific, episodic encoding, frequency estimates for related pairs are less sensitive to actual frequency than are estimates for unrelated pairs.

The next two experiments demonstrate that the conclusions one draws about frequency estimates depend importantly on the relationship between the demands of the final test and the nature of the subjects' encoding strategy. If, for example, the test of frequency judgment is a task which allows subjects to make advantageous use of the association between the members of related pairs, results consistent with a response bias view are obtained.

The fifth experiment extends the findings to judgments of conjoint frequency. Again, the important result is that unrelated pairings show higher frequency discriminability and sensitivity than related pairings. If only low actual correlations are examined, however, results are obtained that look like the operation of a bias. The sixth experiment shows that the results reported in this thesis cannot be attributed to the semantic association between pair members,

(iv)

per se. Instead, the encoding strategy together with the demands of the frequency test are crucial. Finally, the seventh experiment extends the analysis to still other encoding strategies and frequency tests, and confirmation of the main theoretical account is obtained.

In the final section, theoretical issues are re-examined. A union among traditional theories of frequency estimation is proposed. In addition, the theoretical position advocated in this thesis is discussed in the context of more general approaches to human memory.

# Acknowledgements

I thank my friend and advisor, Dr. Ian Begg, who has taught me, by example, the meaning of scholarship. Also this project would not have been completed without the loving support of Emily Harris.

# TABLE OF CONTENTS

	Page
ABSTRACT	iii
ACKNOWLEDGEMENTS	vi
TABLE OF CONTENTS	vii
LIST OF TABLES	ix
LIST OF FIGURES	x
INTRODUCTION	1
FREQUENCY ESTIMATION EXPERIMENTS AND DATA	4
FREQUENCY ESTIMATION RESEARCH	7
THEORIES OF FREQUENCY REPRESENTATION	16
AN EPISODIC/SEMANTIC ENCODING ACCOUNT	19
ILLUSORY CORRELATION	22
AN OVERVIEW OF THE RESEARCH	26
EXPERIMENT 1	28
метнор	30
RESULTS AND DISCUSSION	32
SUMMARY	35
EXPERIMENT 2	37
МЕТНОД	39
RESULTS AND DISCUSSION	41
CONCLUSIONS	44

# TABLE OF CONTENTS

		rage
EXPE	CRIMENT 3	47
	МЕТНОД	50
	RESULTS AND DISCUSSION	51
	CONCLUSIONS	55
EXPE	RIMENT 4	57
	МЕТНОД	59
	RESULTS AND DISCUSSION	59
	CONCLUSIONS	63
SOME	REVIEW AND GENERAL DISCUSSION	66
EXPE	RIMENT 5	70
	метнор	72
	RESULTS AND DISCUSSION	74
	CONCLUSIONS	77
EXPE	CRIMENT 6	79
	метнод	83
	RESULTS	85
	CONCLUSIONS	86
EXPE	CRIMENT 7	89
	МЕТНОД	93
	RESULTS AND DISCUSSION	95
	CONCLUSIONS	108
SUMM	ARY, CONCLUSIONS AND GENERAL DISCUSSION	111
	RELATION TO TRADITIONAL THEORIES OF FREQUENCY ESTIMATION.	116
	OTHER THEORETICAL APPROACHES	121
	PREDICTIONS AND EXTENSIONS	127
REFE	RENCES	130

# LIST OF TABLES

Table	e Numb	er	Following	Page
	1:	Mean error scores for the frequency		
		discrimination task of Experiment 1		33
	2:	Ordinal predictions and results for		
		comparisons of error rates under two		
		hypotheses		34
	3:	Proportion of choices in which the related	•••	
		pair was chosen as having occurred more		
		often in study		42
	4:	Mean error scores for the unequal frequency		
		discrimination test		42
	5:	Ordinal predictions and results for comparison	S .	
	÷	of error rates under two hypotheses for two		
		study instructions		43
	6:	Examples of triplets in the three encoding	,	
		conditions of Experiment 6	8. 	83
	7:	Proportion of related choices for pairs of		
		equal actual frequency		98
	8:	Mean error scores for the frequency		
		discrimination test of Experiment 7		99

(ix)

# LIST OF FIGURES

Figure Num	ber Following	Page
1:	Hypothetical frequency estimation functions	5
2:	Mean frequency estimates for related and unrelated	
	pairs made by subjects who had studied under	
	similarity or contrast instructions	41
3:	Mean proportion of right-hand pair members recalled	
	given left-hand members as cues	51
4:	Mean frequency estimates accompanying cued recall	51
5:	Mean frequency estimates conditionalized on cued recall	52
6:	Cued recall performance by subjects who studied under	
	similarity or contrast instructions	59
7:	Mean frequency estimates accompanying cued recall	
	made by subjects who studied under similarity or	
	contrast instructions	60
8:	Mean frequency estimates conditionalized on cued	
	recall for subjects who studied under similarity or	
	contrast instructions	61
9:	Mean conjoint frequency estimates for subjects who	
	studied under A, general memory; or B, specific	
	correlation instructions	74
10:	Mean frequency estimates for single pair members	
	that came from study pairs on which both members,	
	one member, or neither member had been instances	
	of the orienting category	85

### LIST OF FIGURES (cont'd)

Figure Number

Following Page

11: Mean frequency estimates for pairs in which both members, one member, or neither member had been instances of the orienting category 86 12: 95 Mean on-line frequency estimates for 30 subjects 13: Mean frequency estimates for individual pair members made by subjects who had studied by assessing A, the semantic relationship; B, background pair frequency; or C, running frequency 96 Mean frequency estimates for pairs made by subjects 14: who had studied by aassessing A, the semantic relationship; B, background pair frequency; or C, running frequency 97 15: Cued recall performance for subjects who had studied by assessing A, the semantic relationship; B, background pair frequency; or C, running frequency 100 16: Mean frequency estimates accompanying cued recall for subjects who had studied by assessing A, the semantic relationship; B, background pair frequency; 102 or C, running frequency 17: Mean frequency estimates conditionalized on cued recall for subjects who had studied by assessing A, the semantic relationship; B, background pair 103 frequency; or C, running frequency.

(x1)

This thesis is about the decisions people make when they decide how many times an event has occurred before. It is clear that many types of human decision making require knowledge of the frequency of prior events. Psychological research has shown that the representation of simple event frequency can account for more complex behavior. Performance in concept formation tasks (Bourne, Ekstrand, Lovallo, Kellog, Kiew & Yaroush, 1976; Newmann, 1977), judgments of probability (Estes, 1976) and of contingency (Ward & Jenkins, 1965) have been directly related to the frequency of simple events. In addition, the judgment of simple frequency itself is crucial to much of human behavior. For example, a physician making decisions about diagnosis and treatment needs information about the frequency of symptoms and illnesses. Much of this information must come from the physician's own If the diagnostician's memory is subject to biases or experience. inaccuracies, decisions about treatment may not be optimal.

Some laboratory research has shown that such biases or inaccuracies may exist. It has been reported that, when pairings like <u>bacon-eggs</u> and <u>lion-notebook</u> occur many times, people systematically and grossly overestimate the frequency of the related pairings (Chapman, 1967). This overestimation has been attributed to the "relatedness" of the pairs of associated items and results in subjects overestimating the frequency of pairs whose members are highly associated. In terms of the medical example, a symptom may be judged as highly indicative of a particular disease, not necessarily because the two occur together frequently, but because the two (for whatever reason) are associated, or "seem" to go together. This overestimation has been called

"illusory correlation", is characterized as a response bias, and has clear implications for many aspects of human decision making. This phenomenon has important consequences for the confidence we place in the judgment of observers, for theories of decision making, and for techniques one would recommend to improve decisions.

This thesis addresses the issue of how people make judgments about the frequency of related and unrelated events. First, it will be argued that previous work provides some empirical and theoretical grounds to doubt the characterization of illusory correlation. That is, instead of an overall bias towards the overestimation of related events, frequency estimates for related events may be less sensitive to actual frequency than estimates for unrelated events. This thesis also provides direct evidence that frequency estimates for related events show less discriminability (less sensitivity to actual frequency) than estimates for unrelated events. This result applies to the conditions usually used in frequency judgment experiments. In less typical circumstances, the pattern is altered in predictable ways. The finding of discriminability difference occurs in experiments where old/new discrimination is most crucial to frequency estimation. When recall becomes a prerequisite for estimation, results are obtained that are more appropriately described as a bias. As well as changes in test procedure, it will be shown that changes in study instructions yield similar predictable differences in the pattern of frequency estimates.

It will also be argued that differences in the characterization of illusory correlation are related to other ways in which previous experiments were conducted. It is important to note that, in the absence of differences in overall magnitude, insensitive frequency estimates will be closer to the task mean than more sensitive estimates.

It will be shown that, at low levels of actual frequency, related events receive higher estimates than unrelated. It will be further argued, however, that this occurs, not because of a bias, but because related events receive estimates of lower discriminability. It will be shown that the erroneous conclusion, that related events show a bias towards overestimation, may have occurred because previous experiments examined only low actual frequencies.

Theoretical approaches to frequency estimation will also be The issue of how frequency estimates for related and discussed. unrelated events differ has consequences for our understanding of a number of basic characteristics of memory. This thesis will argue that a crucial distinction is one between recognition and recall, and that frequency estimation is more closely associated with (and logically inseparable from) recognition. It will be shown that under typical circumstances, frequency estimates are much less determined by subjects' ability to recall. I will argue that some theories [availability theory (Kahneman & Tversky, 1973), for example] are theories about recall, and, therefore, cause us to seek differences that are described as a bias. Other theories also concentrate on the magnitude of frequency estimates [the frequency theory of verbal discrimination (Ekert & Kanak, 1974), for example) with little regard to sensitivity. The theory proposed in this thesis asserts the crucial role of recognition in frequency judgments and, consequently, leads us to seek differences in the sensitivity or discriminability of frequency estimates.

A number of other issues will be addressed. It will be argued, for example, that frequency information accrues to the entire study event (in this thesis, a word pair). When a frequency test requires

that the subject reinstate the study event, manipulations of item type or study strategy that yield recall differences will also produce parallel differences in frequency estimates. Another issue that will be addressed is utility of memory information. Different tasks may require different sorts of memory information. Optimal performance in any one task requires that earlier study be appropriate for the later test. Thus, there is no absolutely best way to encode or process to-beremembered material. The value of any particular type of memory information lies in the interaction of test requirements and previous processing strategy. A third issue discussed in this thesis is the form of memory information about frequency. A distinction has been drawn between theories that postulate the retention of specific frequency information (sometimes called propositional views), and theories that assert that frequency judgments are derived from more general memory information (sometimes called inferential views). I will argue that both types of encoding occur, and that both are mutually compatible in a single theory. However, before detailed reviews of previous work and the presentation of experimental results, it is necessary to review what happens in frequency estimation experiments and how the data are usually presented.

### Frequency Estimation Experiments and Data

Frequency estimation experiments typically involve two discrete stages. First, subjects study a list of items, each of which occurs a varying number of times throughout the list. The study items are usually individual words but could be anything. In all the experiments reported in this thesis, the study items are word pairs. Some study pairs are related (e.g., dog-cat, high-low) and some are unrelated

(e.g., <u>moose-piano</u>, <u>river-sad</u>). Second, after study, subjects are tested. Often, this test consists of a list of the items that were presented during study. The subjects' task is to indicate, for each item, how many times it occurred during the study phase. Other ways of testing for frequency information include frequency discrimination. In such a test, individual study items are presented in pairs and the subjects' task is to indicate which item occurred more often during the study phase.

The data from frequency judgment experiments can be presented in a number of ways. One way is a graph in which frequency estimates are plotted as a function of actual frequency (See Fig. 1). This method of presentation is usually associated with frequency tests in which subjects give numerical estimates. Because data from other kinds of frequency tests can be thought of in terms of this function, relating judged to true frequency, and because this method of data presentation is easy to understand, it will be used most often in this thesis. Look again at Figure 1. As shown by the dotted line, perfect performance is represented by a straight line with unit slope and an intercept of zero. Typical results (solid line) depart from perfect performance. The least frequent events are overestimated and the most frequent events are underestimated (Begg, 1974).

Frequency estimation experiments often include manipulations of study instructions or item type, for example. In analysis of the effects of experimental variables on frequency estimates, two general patterns are observed. First, experimental treatments can result in parallel functions that differ only in intercept or overall mean estimate (solid and dashed lines of Fig. 1). Second, experimental





ACTUAL FREQUENCY

treatments can result in functions that differ in slope (dotted and solid lines, and dotted and dashed lines of Fig. 1). Slope differences constitute statistical interactions between the effect of actual frequency and the effect of the experimental manipulation on subjects' frequency estimates. Sometimes slope differences are accompanied by mean differences (dotted and dashed lines) and sometimes they are not (dotted and solid lines). By definition, a steeper slope indicates sharper discrimination in the frequency estimates given items at the various levels of actual frequency. In other words, a steeper slope indicates that frequency estimates are more sensitive to actual frequency.

An experiment by Begg (1974) provides a useful example of discriminability differences in a frequency estimation experiment. In this experiment, subjects studied a list of concrete and abstract words. Each word occurred between 1 and 17 times. The frequency estimates made after the study phase produced a pattern somewhat like Figure 1 (dotted=concrete, solid=abstract). Although concrete and abstract items did not differ in median estimate, the estimates for concrete items increased more steeply with actual frequency than did the estimates for abstract items. The steeper slope for concrete items indicates sharper frequency discrimination. These results for concrete and abstract words are very similar to the results for related and unrelated pairs reported As we shall see, related pairs will show a later in this thesis. shallower slope (lower frequency discriminability) without mean differences. As mentioned above, this slope difference may appear to be a mean difference if one considers only the estimates given pairs of lower actual frequencies.

Before turning to reviews of previous empirical and theoretical work in frequency estimation, consider one final point. Not all frequency judgment experiments yield straight lines for the function(s) relating judged to actual frequency. Often these estimation functions are negatively accelerated curves. Under such circumstances, an interaction between the effects of actual frequency and some other experimental variable still constitutes a difference in the discriminability with which subjects make frequency estimates. This issue of discrimination is important. High discriminability (steep slope) of frequency estimates means high sensitivity to actual The average magnitude of frequency estimates need not frequency. reflect memory information about actual frequency. For example, a shallow frequency estimation function of the correct average magnitude could result from some very general memory information that is not sensitive to the actual number of times events have occurred. Again, by definition, a steep slope indicates high discriminability that, in turn, indicates high sensitivity to actual frequency. As mentioned above this thesis concerns discriminability differences in the frequency estimates given related and unrelated events. A brief review will show that previous empirical work has found both magnitude and discriminability differences with other materials.

### Frequency Estimation Research - Slope and Magnitude Factors

This section will examine some of the empirical findings about frequency judgments. The goal of this examination is some understanding of how frequency judgments are accomplished as a basis for making predictions about how frequency judgments for related and unrelated events might differ. This review will show that some experimental

manipulations affect the magnitude of frequency estimates and others affect the discriminability. Some manipulations affect both, while still others affect neither magnitude nor discriminability. There are commonalities within each set of results that may permit predictions about frequency judgements for related and unrelated events. One of the interesting results of many experiments is that frequency judgments are unaffected by a number of manipulations known to affect other measures of memory retention.

The frequency estimation function is much the same despite differences in study instructions about the kind of memory test that will follow. Subjects told to expect a test of frequency produce estimates indistinguishable from estimates produced by subjects given general memory instructions (Hasher & Chromiak, 1977; Underwood, Zimmerman & Freund, 1971) or instructions to expect a test of recall (Howell, 1973) or recognition (Harris, Begg & Mitterer, 1980). Subjects produce much the same frequency estimates whether or not they receive practice with frequency counting or specific feedback about the accuracy of their performance (Hasher & Chromiak, 1977). Frequency estimates are apparently unaffected by the age of subjects (Hasher & Chromiak, 1977), the duration of item exposure (Hintzman, 1970), the meaningfulness of the study items (Williams & Underwood, 1970), and variations in the typescript in which items are presented (Rowe, 1974). The absence of an effect for all these experimental variables is important because other measures of memory retention are affected by these manipulations (Hasher & Zacks, 1979).

Hasher and Zacks (1979) point out that frequency estimates are independent of many other aspects of memory. The authors review

evidence that estimates of duration, recency judgments and recall performance are independent of frequency judgments. Hasher and Zacks conclude from the absence of effects for these variables that the accrual of frequency-of-occurrence information is highly automatic, requiring little energy from a limited capacity attentional store. Indeed, there seems to be convincing evidence that the encoding of frequency information is at least "non-optional" (Hintzman & Stern, 1978). If the argument that frequency information accrues automatically is correct, then we might be tempted to conclude that the frequency estimates for related and unrelated events will not differ. However, as mentioned above, it has been reported that frequency estimates for related and unrelated events do differ (Chapman, 1967; Tversky & Kahneman, 1973), although, as we shall see, there is some reason to question the exact form of the reported difference. Also, despite the reported lack of effect for some variables, several manipulations do affect frequency estimates.

Much of the work on the effects of experimental manipulations on the subjective representation of frequency is a result of research in verbal discrimination learning. The results of this research are often reported in terms of the effect of a manipulation on the magnitude of frequency estimates, and effects on the discriminability of estimates seem to have been largely ignored. The verbal discrimination task requires a subject to study a list of word pairs. The experimenter designates one word from each pair as correct, and the subject's job is to learn which words are correct and to indicate his choice before feedback is provided (Ekert & Kanak, 1974). The dominant theory to explain performance in such a task is frequency theory (Ekstrand,

Wallace & Underwood, 1966). The occurrence of an event, its pronunciation, rehearsal, and implicit associative responses are all presumed to increment the internal representation of its frequency (Ekert & Kanak, 1972). After practice, subjects should perform accurately in the verbal discrimination task, if they choose items that possess the greater subjective frequency. For obvious reasons, then, much research has concentrated on discovering those experimental treatments that alter the magnitude of subjective frequency with little attention to changes in frequency discriminability. Thus, it is sometimes hard to tell whether the absence of a discriminability difference means that it was not obtained or not looked for.

Many of the manipulations reported to have effects on other memory tasks have been investigated with frequency estimation. Thus. Proctor and Ambler (1975) reported that rehearsal strategy affects the magnitude of frequency estimates. Subjects who were instructed to rehearse items freely gave higher frequency estimates than subjects who were only permitted to rehearse the current study item. It has also been reported that distributed practice produces higher frequency estimates than massed practice (Reichardt, Shaughnessy & Zimmerman, 1973; Rose & Rowe, 1976). Also, frequency context appears to affect frequency judgments (Rowe & Rose, 1977); words presented in the context of a list in which they were the lowest frequency items, received higher final frequency estimates than items (of equal actual frequency) that had been presented in a list in which they were the highest frequency words. Leicht (1968) reported that study items that have more verbal associates within the study list receive higher frequency estimates than items with fewer intralist associates. Reichardt et al. (1973) assessed

The independence of situational frequencies by presenting subjects with two study lists in which the second list either contained a completely new set of words, or was composed of the same items as the first list occurring with different frequencies. When subjects were asked to estimate the frequency of the words in the second list, subjects who received a new second list produced estimates equal in magnitude to judgments for the first list. Subjects whose second list was composed of list l items produced higher frequency estimates than subjects who received new second list items.

Several experimenters have reported differences in magnitude that are accompanied by differences in the discriminability of frequency. estimates. The reader should note that many of the manipulations that affect frequency discriminability are also known to affect old/new discriminability (i.e., recognition performance). Thus, Rowe (1974) reported that subjects who studied a list under a semantic orienting task (by focusing on meaning) produced higher and steeper later frequency estimation functions than subjects who studied under a task which focused on sound. Similarly, Rowe (1973) showed that the verbatim repetition of homonyms produced higher and more discriminating later frequency estimates than did repetitions focusing on different meanings (see also Hintzman & Stern, 1978; Rose, 1980). Rowe and Rose (1977) reported that imagery instructions produce higher and more discriminating later frequency estimates than frequency instructions. Pronounced study items produce higher and more discriminating frequency estimates than unpronounced items (Hopkins, Boylan & Lincoln, 1972; Ghatala, Levin & Wilder, 1973; 1975). Johnson, Taylor and Raye (1977) reported that items that are tested often receive higher and more

discriminating frequency estimates than items that are tested less often, even though study frequency is equal.

There are two points to make about this catalog of results. First, despite the apparent automaticity of subjective frequency accrual, several manipulations <u>do</u> reliably affect subjective frequency. By and large, these effective variables are known to affect other kinds of memory performance. Thus, for example, greater rehearsal, distributed practice, appropriate context, deeper processing, and imaginal study all increase memorability as well as increase frequency estimates. The second noteworthy point is that, sometimes, findings are understood in terms of the effect of a manipulation on the magnitude of frequency estimates while affects on discriminability are often ignored. As noted above, the discriminability of frequency estimates is just as important as the magnitude. There are still other manipulations that seem to produce only discriminability differences.

In the following list of results, note that for the most part the same subjects participate in all experimental conditions. This is in contrast to the results reviewed above. Magnitude differences seem to occur when different subjects participate in the various conditions. When the same subjects participate in the conditions of interest, differences in discriminability alone seem more likely. Thus, imagery is one variable that produces slope differences without differences in overall magnitude. Concrete items receive estimates that have a steeper slope than do abstract items but there is no overall magnitude difference (Begg, 1974; see also Goedel & Thomas, 1977).<sup>1</sup> Time is

<sup>1</sup>Abstract nouns receive higher estimates of background frequency when concrete and abstract items are equated for background frequency (Galbraith & Underwood, 1973). The opposite result appears to obtain with judgments of familarity (Begg & Rowe, 1972).

probably the most reliable producer of discriminability differences. As time since study increases, the discriminability of frequency estimates decreases (Begg, 1974; Underwood et al., 1971). It is also reported that subjects retain a reasonably accurate notion of the mean level of actual frequency so that only the slope and not the average magnitude of estimates changes with delay. Finally, consider an experiment by Malmi (1977). Malmi presented a study list under two different contexts. One context consisted of a set of five filler items that occurred six times each and the other context consisted of thirty fillers presented once Items studied under the low frequency context (30x1) yielded a each. steeper slope for the final frequency estimation function than items studied under the high frequency context (5x6) and there was no overall magnitude difference. These results appear to contradict those of Rowe and Rose (1977) who reported a main effect for context and no significant interaction with actual frequency under very similar experimental conditions.

The point to take from this latest list of experimental findings is that the manipulations that increase the discriminability of frequency estimates are, generally speaking, manipulations that are known to increase other measures of memory retention. For example, relatively shallow processing, relatively abstract items, and study-test delay all decrease retention as well as decrease the discriminability of subjective frequency. In summary, then, it appears that, of all the experimental manipulations known to affect memory retention, some do not affect frequency judgments, some affect the magnitude of subjective frequency, some affect the discriminability of frequency judgments, and

some appear to be both magnitude and slope factors. How does this review aid in an understanding of how frequency judgments are accomplished? How will frequency estimates for compound events differ from estimates for simple, single-item events? Are we any closer to knowing what to expect when subjects must estimate the frequency of related and unrelated pairs?

For answers to these questions, consider two major points already mentioned. First, magnitude differences seem much more likely if the experimental design involves a between-subjects manipulation. That is, if different subjects undergo the various levels of the experimental manipulation, differences in the overall magnitude of frequency estimates between the experimental conditions are more likely. Magnitude differences do not seem to occur when the experimental design involves a within-subjects manipulation. As noted by Begg (1974), subjects who study all classes of items, or study under all conditions have the opportunity to notice, during study, the relative frequency with which the classes of items occur. Knowing that there seemed to be about the same number of concrete as abstract items, for example, might allow a subject to adjust later frequency estimates so that both item types received estimates of about the same average magnitude. A subject could do this even if he or she possessed less information about the actual study frequency of the individual abstract items than about the frequency of the concrete. Subjects who participate in only one condition, or who study only one class of items, have no access to information about relative frequency and, therfore, could not adjust their estimates to reflect the overall task mean.

A second important point to take from the above review and from other research (Flexser & Bower, 1975; Harris et al., 1980; Howell, 1973) is that at least some of the variables that affect other measures of memory performance also affect frequency estimates. However, frequency estimates appear to be more closely associated with some measures of retention than with others. There is evidence (Flexser & Bower, 1975; Harris et al., 1980) that frequency judgments and recognition decisions are mutually dependent. Thus, those study items or experimental conditions that produce the most accurate recognition performance can be expected to yield the most discriminating frequency estimates.

Now that these two generalizations have been drawn for simple, single-item events, we can return to the questions about more complex events. Specifically, how will related and unrelated pairs differ in frequency estimation? On the basis of the literature review above, one may speculate that that class of items that is better recognized will show higher frequency discrimination. With single words, at least, it is known that recall of frequent words exceeds recall of rare words (Hall, 1954) while recognition of rare words exceeds recognition of common words (Shephard, 1967). <u>Fruit-apple</u> is probably a more frequent event than <u>piano-dog</u>. The intimate empirical connection between recognition and frequency estimation suggests that unrelated pairs should show steeper slopes for the frequency estimation functions (i.e., greater frequency discriminability), than related pairs.

Should one expect any overall magnitude differences between frequency estimates for related and estimates for unrelated pairs? If the results from single-item experiments can be generalized to compound

events, the answer depends on the experimental design. It seems most reasonable to present both related and unrelated pairs to each subject at study and test. If subjects can adjust their frequency estimates to reflect the relative frequency of whole classes of study items, one may expect no important overall magnitude differences. Thus, the review of a number of experiments employing single words as study and test items leads to the speculation that the important difference between related and unrelated pairs will be greater frequency discriminability for unrelated pairs.

#### Theories of Frequency Representation

Based on known empirical relationships it is possible to speculate about the pattern of results that should emerge when subjects estimate the study frequency of related and unrelated pairs. Considerable attention has been paid to the development of theories of frequency representation. Part of this attention is due to a basic interest in frequency representation for its own sake, and part of the attention results from attempts to explain performance in other tasks in terms of the representation of frequency. The result of this attention is a considerable body of theoretical work on frequency. Therefore, consider theories of frequency judgment in an effort to make a theoretical prediction about estimates for related and unrelated pairs. First, Underwood's (1969) attribute theory postulates that the trace for an item is a bundle of attributes, one of which is specifically sensitive to frequency. In the case of related and unrelated pairs it is unclear as to whether "memory item" should refer to studied pairs or to individual pair members. In either case, however, attribute theory provides no explicit information about how semantic attributes can

affect the accrual and use of the frequency attribute. In fact. Underwood states that those attributes are independent. List marker theory (Anderson & Bower, 1972) assumes that each occurrence of an item establishes a list marker at the permanent address of the item, and frequency judgments require an estimate of the number of list markers. This theory seems to assert that frequency information accrues to individual pair members; it seems unreasonable to assume a "permanent address" for semantically unrelated pairs of words. When subjects must estimate the frequency of intact related and unrelated pairs, list marker theory gives no explicit basis for predicting any difference in the establishment or retention of list markers for the different pair Additional assumptions involving semantically based directional types. tags, for example, are necessary to predict any difference. Multiple trace theory (Hintzman & Block, 1971) states that each occurrence of an event establishes a separate trace and frequency judgments necessitate an estimate of the number of traces. Contrary to tag theory, multiple trace theory seems to assert that frequency information would accrue to studied pairs rather than individual pair members. However, multiple trace theory explicitly states that memory traces are temporally independent. If this is true, multiple trace theory does not predict any difference in the number or discriminability of traces for related and unrelated pairs, based on pre-experimental frequency, at least. Finally, the availability heuristic (Tversky & Kahneman, 1973) asserts frequency judgments do not depend upon frequency-specific that information, but are inferences based on the availability or the ease of retrieving a memorial instance of the test event. Given a set of intact related and unrelated pairs, availability theory would predict higher

frequency estimates for that class of items that subjects can retrieve more easily, but the theory provides no explicit statement of how related and unrelated pairs would differ in availability.

It should be noted, however, that each of the theories discussed above will predict a difference between the frequency judgments for related and unrelated pairs if one class of items has a different probability of being forgotten between study and test. Availability theory (Tversky & Kahneman, 1973), of course, would predict higher estimates for the more available class of items. At first glance, at least, it seems quite reasonable to assume that related pairs, perhaps because of higher frequency as units, would enjoy a retrieval advantage and produce higher frequency estimates than unrelated pairs. Similarly, the other three approaches predict differences in frequency estimates based upon item memorability. Regardless of whether frequency estimates come from inferences based on more general information (Hintzman & Block, 1971) or come from specifically encoded frequency information (Underwood, 1969), all the approaches in their simplest forms, predict a reduction in the slope of the estimation function and a reduction in the magnitude of estimates for that class of items that has a lower probability of being retained from study to test. The loss of attributes, list markers, or individual traces should result in lower and less discriminating frequency estimates. Although not explicit in any of the theories, assumptions can be added to explain the meanpreserving tendency evident in within-subjects experiments (e.g., Begg, 1974). Either the frequency attribute can decay to the task mean (instead of zero), or estimates of list markers or memory traces can include a comparison with a subjective estimate of the task mean, for

example. The problem is that the theories themselves do not make a statement about the relative memorability of related and unrelated pairs. If such a statement were made, predictions about the pattern of results expected in frequency estimation would follow. The next section will provide a theoretical basis for such a statement.

#### An Episodic/Semantic Encoding Account

First of all, the position of this thesis is that the memory task most crucially involved in standard frequency estimation is some kind of old/new decision. Although this old/new decision does not have to be conscious or isomorphic with measures of recognition, the present position is that subjects, faced with a collection of intact related and unrelated pairs, must make some differential response to the pairs dependent on whether they occurred during study. Much less important in standard frequency estimation is any memorial ability associated with being able to generate all or parts of the study pairs. The second assertion of this theory uses the distinction between episodic and semantic memory made by Tulving (1972). Episodic memory is conceived of as a store of events coded in terms of spatial and temporal characteristics, and semantic memory is thought of as a store coded in terms of associative, permanent characteristics. The point of this second assertion is that the old/new discrimination mentioned above, and, consequently, frequency estimation can only be correctly accomplished with reference to episodic information. The third assumption of this account is that subjects typically behave differently towards a related pair than towards an unrelated pair. Subjects typically attend to the semantic characteristics of related pairs. The encoding of a pair like lion-tiger focuses on general categorical

characteristics of the related pair. On the other hand, the typical strategy employed for unrelated pairs involves attention to relatively unique, episodic characteristics. Thus, the remembered information is the memorial representation of the study processing. The encoding of a pair like piano-apple cannot focus on general categorical features and instead must focus on characteristics of the pair more closely associated with its time and place of occurrence. The approach assumes, therefore, that the difference between related and unrelated pairs lies in the quality of episodic information that accrues. Because of different encoding strategies, subjects' remembered information about the unrelated pairs is more specific to the occurrence event, while for related pairs retained information is biased toward permanent characteristics. Therefore, the frequency judgments for unrelated pairs will show more discriminability than the judgments for related pairs. The account states that, at test, old/new discrimination (again, not necessarily conscious) is a prerequisite to frequency estimation. For both kinds of study pairs, subjects first must decide whether an item has occurred before assigning it to a frequency class.

There are two points to mention about this episodic/semantic account. First, the use of the episodic/semantic distinction here need not imply separate, independent memory stores. It does imply merely that the encoding of an item is variable and can involve relatively general, associative, lexical information or can be devoted to temporal, spatial information. Second, this episodic/semantic approach can be thought of as a special case of a more general approach to frequency judgment and other memory performance. That is, any manipulation which produces a difference in background discriminability will result in

changes in situational discrimination. This general approach would argue that any change in the way subjects deal with study items from background to the experiment or between experimental phases should increase situational discrimination on a subsequent test (Underwood & Freund, 1970). This approach can account for the effects of underlining (Radtke, Jacoby & Goedel, 1971), imagery (Wallace, Murphy & Sawyer, 1973; Goedel & Englert, 1978; Rowe & Paivio, 1971) and pronunciation (Ghatala et al., 1973) upon subjective frequency. That is, subjects' typical strategy for words they encounter probably does not entail underlining, forming a mental image or overt pronunciation. Manipulations that require a subject to perform those atypical activities could yield more discriminating recognition and frequency estimation by differentially improving the episodic memory information. Because the experiments reported here only involve manipulation of subjects' episodic/semantic encoding for related and unrelated pairs, further discussion of the more general approach and of other theories about the representation of subjective frequency will be saved for the general discussion section.

It follows from the statements above that unrelated pairs will be better discriminated from background in standard frequency estimation. The statement that unrelated pairs are better remembered may seem counterintuitive but this statement refers only to measures of recognition. No doubt, measures of free and cued recall would show opposite results. One of the main points of this thesis is that a conclusion, about what class of items is better remembered, depends on the memory test used. Because the standard situation involves a withinsubject manipulation of pair-type, better memory for unrelated pairs will result in better discrimination across levels of actual frequency for unrelated pairs. Estimation functions should be steeper for unrelated than for related pairs. The theory presented above and the previous analysis of empirical relationships lead to the same prediction. There is a problem, however, because some research with subjects' judgments of correlation has led to quite opposite conclusions. These results will be examined next.

### Illusory Correlation

In a study designed to examine illusory correlation, Chapman (1967) presented pairs of items such that any one of four items appeared on the left (e.g., boat, lion, bacon, blossoms) and any one of three items on the right (e.g., tiger, eggs, notebook). Each of the 12 possible pairings occurred an equal number of times (48, 120, or 240). In the test, subjects were required to state the proportion of times that, given a left-hand member, a particular right-hand member followed it. The correct answer is always 33%. However, subjects systematically overestimated (by about 10% on the first test), the conjoint frequency of related items.

Chapman's (1967) interpretation of this result is that the associative connection between the two events improves retention of related pairs. The finding that the correlation between related items is overestimated has been extended to situations involving clinical judgment (Chapman & Chapman, 1969; Golding & Rorer, 1972). A variety of manipulations have been employed, and researchers have reported a strong "[bias] resistant to change even under intensive training conditions (Golding & Rorer, 1972; p. 258)." Subjects consistently reported high conjoint frequencies for expected relationships even when
the actual correlation was zero or negative, and they tended to ignore unexpected relationships. For example, in correlation judgments involving Rorschach responses and psychiatric diagnoses, clinicians overestimated the situational correlation of responses like "female buttocks" with the diagnosis of homosexuality when the two events were, in fact, uncorrelated. Presumably, because of the face validity of the connection, subjects show a bias toward overestimating the correlation between such items even though the situational and clinical relationship is independent.

Tversky and Kahneman refer to the work on illusory correlation in their discussion of the availability heuristic (Tversky & Kahneman, 1973). The availability heuristic is defined as the evaluation of the frequency of classes or the probability of events by the ease with which relevant instances come to mind (Tversky & Kahneman, 1973, p. 207). The authors demonstrate in several experiments that judgments of frequency and probability closely parallel recall performance. In most of their experiments, subjects are not presented with the items for frequency estimation following a study phase. Instead, subjects are usually asked to estimate the frequency of some class of items, and the estimates clearly show a bias towards higher frequency judgments for more easily retrieved items. In discussing illusory correlation and frequency estimates for related and unrelated pairs, Tversky and Kahneman state that subjects base frequency estimates on the strength of the associative bond between pair members. "Thus when a person finds the association between items is strong, he is likely to conclude that they have been frequently paired in his recent experience (Tversky & Kahneman, 1973, p. 224)." It is not clear that basing frequency

judgments on an assessment of "associative strength" is the same heuristic as judgments based on the ease of retrieval. While it seems quite likely that subjects would retrieve instances of a class of events presented for frequency estimation, it is not so obvious that subjects engage in retrieval assessment when intact study items are tested for f requency. It is my contention that the more important assessment subjects make is the old/new discrimination discussed above. Although Tversky and Kahneman present some data supporting their contention that frequency estimates reflect an assessment of the "associative strength" between pair members, their experiment may not address the issue at The experiment employed pairs of personality traits like alerthand. witty and eager-careful that differed in pilot subjects' assessment of the likelihood that both traits would occur in the same person. At some levels of actual frequency, the "related" pairs received higher frequency estimates than the "unrelated" pairs. It can be argued, however, that both pair types are about equally related in a general semantic sense and that the differences in frequency estimates are due to some factor other than retreivability or associative strength.

Consider two important and related points about the availability hypothesis as applied to the frequency estimates for related and unrelated events. First, much of this research has examined only relatively low actual frequencies and correlations. It is important to realize that, if frequency estimation functions for two classes of items differ in slope and not in magnitude, estimates for the class with the lower slope will exceed estimates for items with a steeper slope at low actual frequencies. That is, a discriminability (or slope) difference will look like a magnitude difference if only low frequencies are considered. In fact, if only one actual frequency is tested, differences in frequency discrimination <u>cannot</u> be observed. In such restricted circumstances <u>any</u> difference in frequency estimates for related and unrelated pairs would probably be described as a bias. The research reported in this thesis will examine a wider range of actual frequencies to permit observation of discriminability differences as well as differences in overall magnitude.

A second point may also explain why the research on illusory correlation has not yielded expected discriminability differences. Much of this work is based, implicitly or explicitly, on the availability heuristic. The availability heuristic is basically a theory about the <u>magnitude</u> of frequency and probability estimates. One class of events receives higher estimates because its members are easier to retrieve. Availability theory contains no explicit statement about the discriminability of such estimates. It is fair to infer, I think, that if frequency estimates come strictly from availability, the discriminability or slope of the recall function should determine the discriminability of the resulting frequency estimation function. Later experiments in this thesis will examine this contention. In any case, the work on illusory correlation may not have yielded discriminability differences because the researchers' theory did not lead them to look for such differences.

#### An Overview of the Research

The research reported in this thesis has several purposes. First, it is necessary to establish the true relationship between frequency estimates and actual frequency for related and unrelated word pairs. This involves examination of a wide range of actual frequencies.

Second, a number of measures of frequency judgment will be examined. Third, subjects will study the experimental items under a variety of orienting tasks. Further, the examination of frequency estimation will be extended to assessment of conjoint frequency or correlation between pair members. This research will attempt to discover how conclusions drawn about this "true relationship" depend upon how subjects deal with study material and how frequency judgments are measured.

The first experiments will deal with the question of the true relationship between actual frequency and estimated frequency. Although there seems to be some evidence that subjects will show a bias towards the relative overestimation of situational frequency of related pairs, analysis of other empirical relationships and the theory proposed in this thesis argue for a different relationship. Instead of a simple blas, this position predicts that the important difference between frequency estimates for related and unrelated pairs lies in the discrimination with which subjects make such judgments. That is, unrelated pairs should show more discriminability, (steeper slope) for the function relating judged to true frequency. This difference in discriminability occurs because, during study, subjects attend to the associative relations within pairs. This study strategy results in a relatively semantic encoding for related events that emphasizes permanent characteristics, and a unique episodic encoding for unrelated events that emphasizes information specific to time and place of occurrence. Because only episodic information is useful for accurate frequency judgments, unrelated pairs will show better frequency discrimination than will related pairs. In contrast, no important difference in the overall magnitude of frequency estimates is expected

because these experiments involve a within-subjects manipulation of pair type. Again, these are the predictions for the standard frequency test in which intact study pairs are presented for judgment. Other test situations could produce other patterns of results and later experiments will examine this possibility. Subjects who must generate the study pairs given one member as a cue, should produce frequency estimates that depend on recall. Related pairs are much more likely to be generated and would consequently show a pattern of frequency estimates more like that expected by the availability hypothesis. If frequency information accrues to the entire study event, frequency estimates for individual pair members should show the same pattern. Finally, manipulations of study instructions should alter the pattern of frequency estimates by altering the study strategy subjects employ. Strategies that emphasize the episodic characteristics of related pairs, for example, should remove the discriminability differences predicted above.

## Experiment 1

The purpose of this experiment is to find out whether differences in frequency discriminability exist. In Experiment 1, pairs of words were presented two, four, or eight times in a study list. There were equal numbers of related and unrelated pairs at each level of study frequency. This experiment employed a frequency discrimination test, in which subjects were not required to make numerical estimates of frequency, because Underwood (1972) has shown that such measures are highly sensitive to differences in frequency judgment. The relative frequency test involved comparisons of two critical pairs (including unpresented pairs) in all possible combinations of pair type and actual frequency. The subjects' task was to indicate which of the two pairs occurred more often in the study list.

This frequency discrimination task provides a means to discriminate between two empirical alternatives. Based on these measures, it should be possible to ascertain whether there is a general bias towards the relative overestimation of related pairs, or whether the difference between related and unrelated pairs lies in a difference in the discrimination or sensitivity with which subjects can make frequency judgments about them. The bias prediction is made by the availability-type argument that asserts that subjects make frequency judgments based on the "strength-of-association" between pair members. The prediction of a difference in discriminability is made by the theoretical position advocated by this thesis. That is, subjects attend to the associative connection between pair members, and in the case of related pairs this attention results in a relatively general, semantic memory trace, while in the case of unrelated pairs, attention to the associative connection results in a relatively episodic representation. Differences in discriminability occur because relatively episodic encodings are more useful for judgments of situational frequency than are semantic encodings.

The specific predictions made by each position are quite different. The experiments by Kahneman and Tversky (1973) that demonstrate the operation of the availability heuristic usually employed a frequency discrimination measure. In such a procedure, subjects compared two classes of items or events and were expected to indicate which class occurred more often. If there is a simple, overall bias favoring related pairs in this experiment, one would expect the following pattern of results. When subjects must choose between pairs of equal frequency, one of which is related and the other unrelated, they should tend, for all levels of actual frequency, to choose the related pair. Even when pairs are of unequal actual frequency, related pairs should be chosen more often.

In addition, the view that there is a bias towards the relative overestimation of related pairs would predict that more errors would occur when the lower frequency pair is related and the higher frequency pair is unrelated. A bias view would make this prediction because lower frequency related pairs would have a frequency representation that was too high and this would produce errors when such pairs were compared with unrelated pairs. There is no reason, from this simple bias view, to expect different error rates (choices of a pair whose actual frequency is lower) for situations in which both pairs are related and situations in which both are unrelated.

On the other hand, the view that there are differences in discriminability between related and unrelated pairs makes different predictions. For comparisons of pairs of equal actual frequency, overrepresentation of related pairs in subjects' choices at lower actual frequencies and/or over-representation of unrelated pairs in choices at higher actual frequencies would indicate a discriminability advantage for unrelated pairs. This pattern would occur because, in general, the frequency representation for related pairs would be closer to the task If subjects have more discriminating information about the mean. frequency of unrelated pairs than they have for related pairs, when an unrelated and a related pair of different actual frequency are compared, subjects should erroneously choose a related pair only at lower levels of actual frequency and erroneously choose an unrelated pair only at higher levels of actual frequency. Again, this pattern should emerge if there is sharper discrimination for unrelated pairs because, in general, related pairs would have a frequency representation closer to the task Finally, there should be more errors for comparisons involving mean. two related pairs than for comparisons involving two unrelated pairs. Method

<u>Subjects</u>. Eighty-five introductory psychology students participated in partial fulfillment of a course requirement.

<u>Materials and Procedure</u>. One hundred pairs were chosen from the Palermo and Jenkins (1964) norms such that the second member of each pair was the highest verbal associate of the first. Twenty of the pairs were designated fillers and the remaining 80 were the critical items for the experiment. Forty pairs were chosen at random to remain intact while the members of the other 40 were randomly reassigned to new partners. Of each group of 40 pairs, 10 were chosen not to occur in the study list, 10 to occur twice, 10 to occur 4 times, and 10 to occur 8 times. The filler pairs occupied approximately the first and last 12 positions in the list. The 310-item study list was videotaped at a rate of 5s/item with a randomly chosen spacing of from 9 to 21 items between repetitions (mean = 15.4).

During list presentation, all subjects performed an on-line frequency estimation task in which, for each pair presented, the subject was to indicate the number of previous occurrences for that pair. After study, every subject completed 1 of 5 forms of a final frequency discrimination test. This frequency discrimination test presented sets of 2 pairs. For each set of 2 pairs, the subject was instructed to circle the pair that had occurred more often in the study list. These frequency comparisons involved pairs of all possible combinations of study frequency (0,2,4,8) and item type (related/unrelated). These combinations of frequency and type resulted in 24 different comparisons (0 related versus 0 unrelated, 0 related versus 2 related, 0 related versus 2 unrelated, etc.). Each test also included two comparisons for each of the equal frequency/different item type combination (e.g., 2 related versus 2 unrelated). Each of the critical pairs of the experiment appeared, at most, once on each form of this final test, and on each form of the test a critical pair appeared in a different frequency comparison. Left/right position was also counterbalanced on the 5 forms of the final test. Each of the 5 forms was completed by 17 subjects, and all subjects were told to circle the one pair from each set that had occurred more often in the study list and to guess if they were unsure. The entire procedure took about 45 min.

#### Results and Discussion

Equal frequency discrimination. In this section, as in all results sections of the thesis, standard deviations appear in parentheses,  $\underline{\alpha} = .05$ , and all results reported as reliable meet or exceed this criterion for statistical significance. The first analysis was performed on the data that came from comparisons involving equal actual frequencies and different pair types (related vs unrelated). By assuming a binomial process (and .5 probability of choosing the related pair) for the 170 comparisons at each frequency level, it is possible to assess the probability of a result as extreme as that obtained. For unstudied items, significantly more of the comparisons (64%) resulted in the choice of the related pair than expected. For twice presented pairs significantly <u>fewer</u> of the comparisons (34%) resulted in the choice of the related pairs presented 4 and 8 times, the choice of related pairs was not extreme (50% and 46%, respectively).

This pattern is not the result expected by a view which states there is a bias towards the overestimation of related pairs. A bias should cause subjects to choose related pairs at all levels of actual frequency. Instead, subjects tended to choose related pairs only at the lowest level of actual frequency. For the other levels of actual frequency, subjects tended to choose unrelated pairs as having occurred more often during study. This is like the pattern expected by a view which states that subjects have more discriminating memory information about the frequency of unrelated pairs. That is, related pairs appear to have a frequency representation closer to the task mean than do unrelated pairs. Different frequency discrimination. The next analysis concerns those comparisons between pairs of different actual frequency. Each subject performed 24 such comparisons, and the comparisons conform to a  $2 \times 2$  design in which the two repeated factors are whether or not the higher frequency pair in the comparison is related, and whether or not the lower frequency pair is related. Mean error scores (out of 6) are shown in Table 1.

The data summarized in Table 1 were subjected to an analysis of variance. It is clear from the table that more errors were made when the higher frequency pair in the comparison was related [means = .69, vs. .48; F(1,84) = 17.57; MSe = .22]. Although these data also suggest an interaction such that homogeneous pairs produce more errors than heterogeneous pairs, the interaction was not statistically significant. In any case, this is not the result that would be expected if there were a bias favoring a general overestimation of the frequency of related pairs. If such a bias were operating, one would expect more errors when the lower frequency pair in the comparison was related. That is, a bias should make related pairs seem as if they have occurred more often than they actually have. The data in Table 1 also do not clearly support a differential discriminability position. Thus, a finer grained analysis was conducted.

<u>Specific predictions</u>. This more detailed analysis examined unequal-frequency comparison data. By assuming either a steeper discrimination function for one pair type or an overall bias for one pair type, it is possible to make predictions about differences in error rates for many of the comparisons. For example, a view that assumes a bias for the overestimation of related pairs should predict that more

Table 1: Mean error scores out of 6 for the frequency discrimination task of Experiment 1 (s.d.).

Higher Frequency Pair in the Comparison

		R	U
Lower Frequency Pair	R	.72 (.84)	.40(.63)
in the Comparison	U	.66 (.75)	.55(.73)

errors occur when subjects must decide between an unrelated item presented 8 times (8U) and a related item presented 4 times (4R) than when subjects must say whether a related item presented 8 times (8R) occurred more often than an unrelated pair presented 4 times (4U). In other words a general bias view would predict that the 8U4R comparison would produce more errors than the 8R4U comparison. On the other hand, a view which assumes a steeper frequency estimation function for unrelated pairs predicts the opposite. Steeper and generally more sensitive representation of frequency for unrelated pairs should result in more errors for the 8R4U than for the 8U4R comparison. For this experiment there are 6 comparisons that can be made for error rates within each level of frequency. There are 6 levels of frequency for comparisons (0/2, 0/4, 0/8, 2/4, 2/8, 4/8), and this results in 36 possible meaningful comparisons of error rates. Of these 36, the biasfor-related-pairs view makes directional predictions for 30. The view that assumes sharper discrimination across frequency for unrelated pairs makes directional predictions in 34 comparisons. The comparisons, predictions and data for this experiment are shown in Table 2. For the bias view, 11 of the 30 predictions are in the expected direction while the differential accuracy predictions result in 22 of 34 in the expected direction. Finally, of the 36 comparisons, 15 involve opposite predictions for the two views (shown by an asterisk in Table 2). Of these 15 discriminating predictions, 11 are in the direction expected by the differential accuracy view while only 3 favor the bias interpretation. This result corresponds to a significant difference by a two-tailed sign test where a = .06.

Comparison A vs. B <sup>1</sup>	Bias Prediction <sup>2</sup>	Diff. Disc. Prediction <sup>3</sup>	Obtai Resu	lned ult <sup>4</sup>
2ROR 2UOU	-	>	.20	.04
2ROR 2ROU	>	>	.20	.11
2ROR 2UOR	<	<	.20	.07
2UOR 2UOU	>	>	.07	.04
2UOR 2ROU	>	-	.07	.11
2ROU 2UOU	- <	<	.11	.04
4ROR 4UOU		>	.06	.01
4ROR 4ROU	> •	>	.06	.05
4ROR 4UOR*	<	. >	.06	.06
4UOR 4UOU	>	>	.06	.01
4UOR 4ROU	>	>	.06	.05
4ROU 4UOU*	<	>	.05	.01
זוחוופ מחמפ	2 <u>1</u>	<u>`</u>	02	.01
		Ś	.02	0
SROR SHOR*	2	Ś	.02	õ
SHOR SHOLL		Ś	0	.01
SUOR SROU	Ś	-	õ	Õ
SPOIL SHOUX	1	>	0	.01
CROO COOC			•	
4R2R 4U2U		>	.25	.29
4R2R 4R2U	>	>	.25	.19
4R2R 4U2R*	<	>	.25	.15
4U2R 4U2U	>	>	.15	.29
4U2R 4R2U	>	_	.15	.i9
4R2U 4U2U*	<	>	.19	.29
0-			01	01
8R2R 8020		2	.01	.01
SKZR SRZU	2	2	.01	.05
OKZK OUZK*			.01	.04
SUZR SUZU	2	1	.04	.01
OUZK SR2U*	2		.04	.05
	(		•UJ	• U L

Table 2: Experiment 1 - Ordinal predictions and results for comparisons of error rates under two hypotheses.

cont'd

#### Table 2 (cont'd)

8R4R	8040		-		>	.18	.16
8R4R	8R4U*		>		<	.18	.25
8R4R	8U4R*		<		>	.18	.07
8U4R	8040*		>	-	<	.07	.16
8U4R	8R4U*		>		<	.07	.25
8R4U	8U4U*		<		>	.25	.16

<sup>1</sup>Pairwise comparisons of error rates <sup>-</sup> for example, the first comparison, 2ROR vs. 2UOU, involves the comparison of the probability of an error, when subjects must decide which of a twice presented related pair (2R) and an unpresented related pair (OR) occurred more often during study, with the probability of an error when subjects decide between a twice presented unrelated pair 2U and an unpresented unrelated pair (OU).

- <sup>2</sup>Ordinal prediction under the assumption that the <u>only</u> difference between related and unrelated pairs is that of an overall bias towards overestimating the frequency of related pairs.
- <sup>3</sup>Ordinal prediction under the assumption that the <u>only</u> difference between pair types is greater discriminability of frequency information for unrelated pairs.

<sup>4</sup>Probability of an error for A and B, respectively.

\*Comparisons involving opposite predictions.

#### Summary

It is clear that the results of this experiment do not support the notion that there is a simple bias favoring related pairs. Across the experiment, related pairs were not chosen more often than unrelated (1328 times versus 1392). When actual frequency was equal there was no overall tendency for related pairs to be chosen as having occurred more often during study.

On the other hand, the two classes of items seem to differ in the sensitivity with which subjects can make decisions about frequency. It is harder to decide which of two related pairs has occurred more often than which of two unrelated pairs has. For mixed comparisons, subjects make more errors when the higher frequency pair in the comparison is related, especially at higher levels of actual frequency. Finally, for equal frequency comparisons, related pairs are overrepresented in the choices involving unpresented items, while unrelated pairs are overrepresented in the choices involving twice presented items. These results suggest that, instead of a simple bias causing the relative overestimation of related pairs, associative connections operate to reduce the accuracy with which subjects can discriminate between related pairs on the basis of their frequency of occurrence.

In theoretical terms, subjects have information about the frequency of the study pairs that is of poorer quality for related pairs than it is for unrelated pairs. This difference in the quality of memory information is a result of the encoding process subjects employ during study. Because their members have occurred together before the experiment, subjects tend to encode the related pairs in terms of

permanent, associative or semantic characteristics. Because their members are likely to be occurring together for the first time, subjects encode the unrelated pairs in terms of the time and place of the pairs' occurrence. This results in memory information that is unique to the study episode and more useful in a later frequency test. Clearly, information about permanent characteristics of the study pairs would be of little utility in a frequency discrimination test.

## Experiment 2

The purpose of this experiment is to test, via the manipulation of study instructions, one of the theoretical assumptions discussed earlier. This assumption is that subjects deal with related and unrelated pairs differently during study. Subjects attend to the semantic relationship between pair members. Even when instructed to attend to frequency, subjects' typical study strategy is to look for an associative connection between pair members. If these assertions are true, instructions to find similarities between pair members should duplicate the typical strategy. Attention to similarities for related pairs tends to produce a memory encoding that is focused upon permanent semantic characteristics of the pair. Attention to similarities for unrelated pairs tends to produce a memory encoding focused upon the time and place of the pair's occurrence because those are the most obvious shared characteristics.

What should happen if subjects are instructed to look for differences between pair members? Attention to differences for related pairs should require some shift from the typical strategy. Although focused on permanent characteristics, the resulting memory encoding should be different from that produced in encounters with the pair under other instructions. Seeking differences between members of unrelated pairs should allow subjects to focus on relatively permanent characteristics of those items. Following instructions to look for differences, therefore, subjects should have memory information for the two kinds of pairs that is very similar - in both cases focused on relatively permanent, semantic characteristics. According to the

account discussed above, removing differences in encoding strategy should remove differences in frequency discriminability. It is possible that, under instructions to look for differences, the encoding for related pairs would be focused even more upon the time and place of occurrence than the encoding of unrelated pairs. This possibility remains open to question, but the expectation of this experiment that a change in study instructions should at least remove (and possibly reverse) the slope difference between the frequency estimates for unrelated and related pairs.

Begg (1978) used <u>similarity</u> and difference or <u>contrast</u> instructions in assessing memory for sets of related and unrelated items. The results of his experiments and his theoretical interpretation are consistent with the predictions made above. His research showed that <u>contrastive</u> processing (discovery of distinctive features) aids memory for related items. Thus, related pairs were better recognized after <u>contrastive</u> than after <u>similarity</u> processing. Because his experiments addressed different issues, Begg did not include recognition measures for unrelated items. The theory proposed above asserts that recognition of unrelated pairs should be better after similarity processing than after contrastive.

In the second experiment of this thesis, related and unrelated pairs were presented 1, 2, or 3 times each. Subjects were instructed to study by either <u>similarity</u> or <u>contrastive</u> processing. The expectations are simple. First, <u>similarity</u> processing should simulate what subjects usually do in such an experiment. As found in Experiment 1, there should be no overall magnitude difference in the frequency judgments for related and unrelated pairs. Instead of a bias, frequency judgments for

related pairs should show a shallower slope and reduced frequency discriminability compared to judgments for unrelated pairs. <u>Contrastive</u> processing, on the other hand, should alter the typical or usual pattern. By focusing upon the permanent, semantic characteristics of both types of study pair, <u>contrastive</u> processing should remove the difference quality of the frequency information. Thus, <u>contrastive</u> processing should remove the discriminability (or slope) difference obtained under <u>similarity</u> instructions. The theory presented above predicts no bias or overall magnitude difference. However, Begg (1978) found very large differences in recall probability between related and unrelated items after <u>contrastive</u> processing (e.g., .97 vs. .32, Experiment 1). Such large differences in recall could be reflected in frequency judgments by producing a bias towards the relative overestimation of related pairs.

In sum, the episodic/semantic account assumes that the difference in frequency discriminability between related and unrelated pairs results from a difference in the way subjects typically deal with the two pair types. This typical study strategy will be simulated by <u>similarity</u> processing. By removing the difference in the way subjects typically deal with the pairs, <u>contrastive</u> processing should remove the discriminability difference.

#### Method

<u>Subjects</u>. One hundred and fifty-two introductory Psychology students participated, 130 in the frequency discrimination test (68 in <u>similarity</u> and 62 in <u>contrast</u>) and 22 in frequency estimation (11 in each instructional group).

Materials. The same 100 pairs from Palermo and Jenkins (1964) used in Experiment 1 were divided into 20 filler and 80 critical pairs. A new randomly chosen half of the critical pairs remained intact and the members of the remaining 40 were randomly reassigned to new partners. Again 10 of each set of 40 were selected at random to occur 0, 1, 2, or 3 times each in the presentation list. Fillers occupied approximately the first and last 12 positions in the 162 item list. The study list was videotaped at a rate of 10s/item with a mean spacing of 12 items between repetitions.

<u>Procedure</u>. Subjects received one of two sets of study instructions. Subjects in the <u>similarity</u> condition were instructed to find, for each pair, a characteristic the members shared, a way in which members were alike, or could go together in a sentence. Examples were discussed and subjects were told to write a word or phrase summarizing the discovered similarity on a response sheet. The <u>contrast</u> instructions were identical except that subjects were told to find some way in which the members of the pair contrasted or were dissimilar. All subjects were told that a memory test would follow the list.

Subjects completed one of 2 sorts of memory tests. First, the frequency discrimination test consisted of 5 forms constructed by the same method as those of Experiment 1. The only difference was that sixteen comparisons involving pairs of equal actual frequency were given to each subject. Each of these comparisons of equal frequencies involved one related and one unrelated pair. There were also 24 comparisons involving all possible combinations of unequal frequencies and pair types. At least ten subjects completed each form following each study instruction. The remaining 22 subjects completed the

frequency estimation test. For this test all 80 critical pairs were presented in random order and the subjects were told to indicate, for each pair, the actual presentation frequency using the digits 0, 1, 2 or 3. The entire procedure took about 45 min.

#### Results and Discussion

Because a number of measures were employed, results are reported separately for each measure. Absolute estimates will be discussed first followed by the frequency discrimination tests. Next, a finer grained analysis is discussed followed by a comparison among the various measures.

Absolute estimates. The first set of analyses concern the frequency estimates for pairs, shown in Figure 2. These data show a difference in the effect of the instructions on the estimates. Under similarity instructions, the expected pattern was obtained. There is no overall advantage for related pairs but there is the suggestion of the interaction of pair type with actual frequency that would indicate better frequency discrimination and accuracy for unrelated pairs. Subjects who studied under contrast instructions show a different pattern of results. There is no real difference in the slope of the function for related and unrelated pairs; but, overall, related pairs receive higher estimates than unrelated. These conclusions are partly supported by analysis of variance on the subjects' mean frequency estimates, with study instructions a between-subjects factor, and actual frequency and item type within-subjects factors. The analysis yielded a main effect for item type, indicating higher estimates for related pairs [means = 1.44 vs. 1.33, F(1,20) = 12.17, MSE = .040], and the usual maineffect for frequency [F(3,60) = 1268, MSE = .036]. The only reliable

Figure 2: Mean frequency estimates for related and unrelated pairs made by subjects who had studied under <u>similarity</u> or <u>contrast</u> instructions



interaction was between instructions and pair type [F(1,20) = 5.53], MSE = .040]. Although expected, the three-way interaction was not significant. The conclusions are also supported by the results of other measures, however.

Equal frequency discrimination. The analyses for equal and unequal frequency discrimination employed only 100 of the 130 subjects. Subjects were eliminated at random in order to permit analyses of variance on groups of equal size.

The next set of analyses concerned the frequency discrimination tests and the results partly supported the conclusions drawn above. The data for the comparisons involving pairs of equal actual frequency and different type are shown in Table 3. This is clearly not the pattern of results obtained in Experiment 1. These comparisons of pairs of equal actual frequency show an overestimation of related pairs especially for items studied under <u>contrastive</u> instructions. The zero items were not actually studied, of course, and consequently related items are apparently less often selected perhaps because lack of <u>contrastive</u> study results in a frequency representation that is easily discriminated from other pairs. In any case, these results suggest the existence of a bias for the overestimation of the frequency of related pairs, rather than a difference in discriminability.

Unequal frequency. The results for the comparisons of unequal frequency are shown in Table 4. This table shows mean error scores out of 6 for 100 subjects. These data were subjected to a one-way analysis of variance [F(7,392) = 2.21, MSW = .495] to permit a number of planned comparisons. First of all, as is clear from Table 4, <u>contrast</u> instructions resulted in more errors than did similarity instructions

Table 3:	Proportion	of	choices	in	which	the	related	pair	was	chosen	as
	having occu	rre	d more o	ofte	en in s	study	7.				

STUDY INSTRUCTIONS		Frequency				
		0	1	2	3	
Similarity ( <mark>n</mark> = 272)		.56	.61*	.55	•48	
Contrast (n = 248)	,	.39*	.61*	.63*	•56*	

\*Indicates results that differ from chance.

# Table 4: Mean error scores out of 6 for the unequal frequency discrimination test of Experiment 2.

	S	imila	arity I	nstructions	Contrast Inst	ructions	
			Higher	Frequency Pair	in the Compar	ison	
		•	R	U	R	U	
Lower Frequency	R	.58	(.79)	.40 (.54)	.64 (.79)	.80 (.93)	
Comparison	U	.36	, (.57)	.46 (.65)	.42 (.61)	.58 (.64)	

(critical difference = .138, obtained difference = .160). The next planned contrast shows that the data for contrast instructions conform closely to the predictions of a bias for related pairs. This contrast shows that, for contrast instructions alone, more errors occur when the lower frequency pair in the comparison is related (critical difference = .195, obtained difference = .220). This is a prediction that follows directly from a view which asserts that there is an overrepresentation of the frequency of related pairs. The most errors occur when the lower frequency pair is related and the higher frequency pair unrelated (.80 The pattern for similarity instructions is quite different. of 6). Although no other planned contrasts resulted in significant differences, this pattern of results is very similar to that of Experiment 1 in which heterogeneous comparisons resulted in fewer errors than homogeneous comparisons, and more errors occurred when the higher frequency pair of the comparison was related. However, in neither experiment was the difference significant.

Specific predictions. Finally, a finer-grained analysis similar to that conducted in Experiment 1 was conducted. The comparisons, predictions and obtained error rates are shown in Table 5. Because error rates are even lower in this experiment, even less discriminnation among positions is possible. For <u>similarity</u> instructions, 14 of 32 possible comparisons favor the position that the only difference between related and unrelated pairs is one of greater discrimination across frequency for unrelated pairs. Of 30 comparisons, 13 favor the view that there is a bias in favor of related pairs. For predictions that discriminate the two positions, 5 of 13 favor differential sensitivity

Table 5:

Experiment 2 - Ordinal predictions and results for comparisons of error rates under two hypotheses for two study instructions.

Comparison A vs. B	Bias Pred.	Diff. Disc. Pred.	Obtained Re Similarity	contrast
1ROR 1UOU 1ROR 1ROU 1ROR 1UOR 1UOR 1UOU 1UOR 1ROU 1ROU 1UOU	- - - - - - - - - - - - - - - - - - -	> < > > <	.14 .02   .14 .04   .14 .04   .04 .02   .04 .04   .04 .04   .04 .04	.16 .11   .16 .02   .16 .04   .04 .11   .04 .02   .02 .11
2ROR 2UOU 2ROR 2ROU 2ROR 2UOR* 2UOR 2UOU 2UOR 2ROU 2ROU 2UOU*	- ~ ~ ~ ~	> > > - >	$\begin{array}{cccc} .12 & 0 \\ .12 & .02 \\ .12 & 0 \\ 0 & 0 \\ 0 & .02 \\ .02 & 0 \end{array}$	.05 0 .05 0 .05 0 0 0 0 0 0 0
3ROR3UOU3ROR3ROU3ROR3UOR*3UOR3UOU3UOR3ROU3ROU3UOU*	- > < > > <	> > > - >	$\begin{array}{ccc} 0 & 0 \\ 0 & 0 \\ 0 & .04 \\ .04 & 0 \\ .04 & 0 \\ 0 & 0 \end{array}$	.02 0 .02 0 .02 .05 .05 0 .05 0 0 0
2R1R 2U1U 2R1R 2R1U 2R1R 2U1R* 2U1R 2U1U 2U1R 2R1U 2R1U 2U1U*	- - - - - - - - - - - - - -	> > - >	.12.09.12.05.12.12.12.09.12.05.05.09	.14 .11 .14 .09 .14 .21 .21 .11 .21 .09 .09 .11
3R1R 3U1U 3R1R 3R1U 3R1R 3U1R* 3U1R 3U1U 3U1R 3R1U 3R1U 3U1U*	- - - - - - - - - - - - - -	> > - >	.02 .05 .02 .02 .02 .02 .02 .02 .02 .05 .02 .02 .02 .05	.04 .11 .04 .07 .04 .11 .11 .11 .11 .07 .07 .11
3R2R 3U2U 3R2R 3R2U* 3R2R 3U2R* 3U2R 3U2U* 3U2R 3R2U* 3R2U 3U2U	- - - - - - - - - - - - - -	> < > < > > < > > < > > < > > < > > > >	.37 .40 .37 .30 .37 .21 .21 .40 .21 .30 .30 .40	.23 .37 .23 .28 .23 .37 .37 .37 .37 .28 .28 .37

<sup>1</sup>Probability of an error for A and B, respectively.

\*Comparisons involving opposite predictions.

and 5 favor the bias view. Thus for <u>similarity</u> instructions, this analysis permits no discrimination between the two competing views.

There is somewhat more discrimination possible for the <u>contrast</u> instructions. Of 32 comparisons, 13 favor differential discriminability while 19 favor a bias-for-related interpretation. For predictions that discriminate between the two positions, 8 favor bias and only 4 favor differential accuracy. Although this does not constitute a significant difference by a sign test, in view of the results of other analyses, <u>contrast</u> instructions do produce different results than <u>similarity</u> instructions. The pattern of results for <u>similarity</u> instructions is very much like the pattern obtained in the first experiment in which related pairs seemed to show, not an overall bias, but lowered discriminability on the basis of frequency compared to unrelated pairs. On the other hand, <u>contrast</u> instructions did produce a bias. After <u>contrast</u> instructions the frequency of related pairs is overestimated relative to unrelated pairs.

### Conclusions

The expectations of this experiment are fairly well borne out. The absolute-frequency-estimation procedure discriminates between the study instructions. Although not always statistically significant, <u>similarity</u> processing seemed to produce frequency estimates that were more discriminating among levels of actual frequency for unrelated pairs. There was no overall difference in the magnitude of estimates for related and unrelated pairs indicating that the difference between the two pair types lies not in a bias but in a difference in the quality or sensitivity of frequency information. <u>Contrastive</u> processing, as expected, altered this pattern of results by making the quality of the

episodic information equal for both pair types. Discriminability differences were removed and an overall difference in the magnitude of frequency estimates was obtained. It may be that large differences in the ability to recall (cf. Begg, 1978) do produce frequency estimates that look like the operation of a bias. In any case, this bias does not seem to occur under frequency or similarity instructions. The relative frequency measures also provide some discrimination between the two study groups. In theoretical terms, the subjects who studied the list under similarity instructions focused on the typical semantic characteristics of related pairs and on atypical aspects of unrelated pairs. This study strategy results in better episodic information about the occurrence of unrelated pairs and a pattern of results like that obtained in the first experiment. On the other hand, contrastive instructions improved the quality of episodic information for related pairs, reversing the pattern and producing results in which related pairs do show an overall bias and/or greater discrimination in judgments. Again, the results are consistent with the idea that the utility of information encoded on any trial varies both with encoding strategy and the nature of the test. Encoding typical, semantic information is of less use in subsequent frequency judgment tasks than is the encoding of episodic information that is more sensitive to the time and place of an event's occurrence.

The main conclusion of this experiment is that, in standard frequency estimation, the associative connection between members of related pairs results in worsened background/situational discrimination for these pairs and that this, in turn, results in decreased discrimination for the frequency estimates for related pairs. Much less

important, in standard frequency estimation, is any improvement in recall of related pairs that might be expected to lead to an overall bias. In standard situations, therefore, some kind of episodic old/new discrimination is the crucial memorability factor for frequency judgments. Atypical study strategies remove discriminability differences and produce results that can be described as the operation of a bias. Large differences in recall probability (i.e., availability) may be responsible for this difference in estimate magnitude. Later experiments will examine the role of availability further. The next experiment shows that the use of different frequency tests can provide results consistent with the existence of a bias towards the relative overestimation of related pairs.

#### Experiment 3

The theory developed so far has asserted that old/new discrimination is crucial to standard frequency estimation. That is, subjects first decide whether the test item has occurred before, then produce numerical estimates. Under ordinary circumstances, related pairs show lower frequency discrimination because subjects have less useful old/new information about related pairs. Subjects are influenced by the associative connection between pair members to make less useful (for the purposes of old/new discrimination) encodings. In other words, the obvious fact that the items are related is so salient that is seduces the subject into encoding that relation, ignoring less general, more situation-specific information in the encoding context. There should be circumstances, however, in which the associative connection between pair members would aid memorability and, consequently, aid frequency discrimination.

The point of this third experiment is to demonstrate that the pattern of results obtained in the first two experiments is due to the combination of two factors. Both the nature of the test and the nature of the study strategy determine subjects' performance. This third experiment employs a test different from the usual frequency estimation task. This test was chosen in order to permit the associative connection between pair members to operate to improve performance. The test chosen was intralist cued recall accompanied by frequency estimation. In such a test, the left-hand members of the study pairs are presented. The subject's task is to produce the right-hand members and a frequency estimate for each pair. Cued recall for related pairs

will certainly exceed that for unrelated pairs. The frequency estimates made simultaneously should reflect that memory difference. That is, in such a situation, the ability to recall items (rather than perform an old/new discrimination) would be crucial to the discriminability of frequency estimation. In a sense, this test may add a new stage to the beginning of the frequency estimation process. This new stage requires the generation of the test items before the old/new discrimination and numerical estimation steps. Whereas the first stage in the standard estimation test confers an advantage on the unrelated pairs, this new first stage should confer an advantage on the related pairs. The result of this added advantage should be greater frequency discrimination for the frequency estimates that accompany the cued recall of related pairs.

Although the expectations are quite clear, there are some procedural problems that might prevent us from seeing the expected results. A potentially serious problem concerns the manner in which recall is accomplished. Cued recall of related items could be accomplished by two routes. Subjects could indeed remember something about the pair's occurrences during study and thereby produce the correct response. Additionally, however, subjects could merely generate a verbal associate for the cue, thereby mimicking correct cued recall without remembering anything about the pair's occurrences during study. Presumably, only the first route could produce correct cued recall for unrelated pairs. However, if the subject had no information about a pair's occurrences during study, his/her frequency estimate could only be a "pure" guess. That is, some of the frequency estimates given "correctly" recalled related items could be undiscriminating guesses,

and this would operate against finding the predicted discriminability differences between related and unrelated pairs. However, conditionalizing frequency estimates on cued recall performance would provide some assessment of the importance of this problem. Steeper estimation functions for correctly recalled unrelated items than for correctly recalled related itms would suggest that some of the "correctly" recalled related items are really being produced by the second nonepisodic route.

In addition, as discovered in Experiment 2, very large differences in memorability can be reflected in the magnitude of frequency estimates. Therefore, if cued recall performance for related pairs is much greater than cued recall of unrelated pairs, some overall, parallel difference in the magnitude of frequency estimates would be expected also. That is, large differences in memorability would result in accompanying differences in frequency estimates that would be described as a bias for related pairs.

Finally, as a sidelight, note that the predictions made by the episodic/semantic account are similar to those made by the availability hypothesis (Tversky & Kahneman, 1973). If judgments of frequency are made strictly on the basis of recall-ease or success, frequency estimates should very closely parallel cued-recall performance. The simplest version of the availability heuristic would also assert that discriminability or slope of the cued-recall function would determine the slope of the frequency estimation function. That is, if frequency estimates come strictly from recall, any slope differences in cuedrecall should be maintained in accompanying frequency estimates. The episodic/semantic account makes a slightly different prediction, however. Because frequency information accrues to the entire study pair, subjects have more accurate frequency information about items that are better remembered (regardless of the slope of the cued-recall function). These two positions are not necessarily incompatible. The availability heuristic makes explicit predictions about the magnitude of frequency estimates and predictions about the discriminability of estimates must be inferred. The episodic/semantic account makes explicit predictions about the discriminability (slope) of frequency estimates. Statements about subjects adjusting estimates to maintain a task mean, for example, must be added to account for magnitude differences.

Nevertheless, the main contention of these predictions is that improved memory (caused by the associative connection) is not important in standard frequency estimation tasks because recall of pair members is not important. In such standard situations, differences in the episodic/semantic quality of memory traces dominate the estimation of frequency. On the other hand, when recall of items is required, the improved memorability conferred on related pairs by the associative connection results in frequency estimates that depend upon recall. Method

<u>Subjects</u>. Twelve introductory psychology students participated in partial fulfillment of a course requirement.

<u>Materials and procedure</u>. Materials, study list and study instructions were all identical to those of Experiment 1. The final test consisted of a sheet on which was printed every left-hand member of all 80 critical pairs (including unpresented items). Beside each cue there were 2 blanks. Subjects were instructed to recall the word that
had appeared with the stimulus item during study and write it in the second space. Subjects were instructed to write the frequency with which the pair occurred during study in the first space, using the digits 0 to 8. If any subject gave an estimate of zero, he or she was instructed to "Guess any word for the second blank." The entire procedure took about 50 min.

#### Results and Discussion

Because of the number of measures involved, results will be reported separately for cued recall, followed by frequency estimates unconditionalized and conditionalized upon cued recall. Next are some special considerations regarding the results for the test items that were not studied.

<u>Cued recall</u>. The first set of results is summarized in Figure 3, which shows the number of items correctly recalled. Cued recall improved with study frequency and related pairs produced a higher level of correct cued recall than did unrelated pairs (means = 7.14 vs. 2.38). These conclusions were supported by an analysis of variance in which actual frequency and pair type were repeated factors. The analysis yielded the typical powerful main effect for frequency [F(2,22) = 303], and a main effect for item type [F(1,11) = 21.08, MSE = 2.22]. The interaction of these two factors was not significant.

<u>Unconditionalized frequency estimates</u>. The second analysis concerned frequency estimates produced in this task, shown in Figure 4. Again, an analysis of variance was performed on the mean frequency estimates in which actual frequency and pair type were repeated factors. Paralleling the cued recall results, this analysis yielded a large effect of actual frequency [F(2,22) = 70.06]; a main effect of pair type Figure 3: Mean proportion of right-hand pair members recalled given left-hand members as cues



# Figure 4: Mean frequency estimates accompanying cued

recall



such that related pairs received higher estimates [4.0 vs. 3.26; F(1,11) = 8.26, MSE = .60]; and no interaction.

Conditionalized estimates. Finally, Figure 5 shows frequency estimates conditionalized upon cued recall performance. For related and unrelated pairs, correctly recalled items received higher frequency estimates than incorrect pairs. In both cases, there is much better discrimination on the basis of frequency, shown by a steeper slope for frequency estimates across actual frequency, for correctly remembered pairs than for all pairs in general, and especially incorrectly recalled pairs. In fact, incorrectly recalled pairs show very little increment in frequency estimates with actual frequency. These results parallel those of Harris et al. (1980). In that experiment it was shown that in standard frequency estimation, unrecognized items show no increment in estimated frequency. Mean estimates for related pairs were: 3.10(1.05)vs. 1.93(.97) for twice presented items (Z = -2.29, Wilcoxon signedranks test<sup>2</sup>); 4.56(1.26) vs. 1.95(1.80) for items presented 4 times (Z = -2.58); and 6.1(1.36) vs. 2.90(1.80) for items presented 8 times (Z = -2.80).The results for unrelated pairs were: 4.10(1.81) vs. 2.06(1.42) for 2's (T = 6, n = 7, nonsignificant, Wilcoxon test); 4.81(1.89) vs. 2.47(1.09) for 4's (Z = -2.31); and 7.28(.79) vs.

<sup>&</sup>lt;sup>2</sup>For experiments 3 and 4, it should be noted that, because differing numbers of subjects showed different recall levels with each class of items, some of the values above represent relatively few observations. Many subjects must be eliminated from conditional analyses because they either get perfect recall with related pairs or zero recall with unrelated pairs. For this reason, because the conditional means are quotients of random variables, and because assumptions about homogeneity of variance are violated, the nonparametric Wilcoxon matched-pairs signed-ranks test was used. The Z scores reported refer to the standard normal deviates associated with the obtained value of the Wilcoxon T statistic.

Figure 5: Mean frequency estimates conditionalized on

cued recall



2.84(1.69) for 8's (Z = -3.06). Thus, in almost every case, estimates for recalled items are significantly higher than estimates for unrecalled items.

A final analysis compared the frequency estimates given correctly recalled related pairs to the estimates given correctly recalled unrelated pairs. Although related pairs received higher estimates in the unconditionalized analysis reported above, the opposite result is obtained when estimates are conditionalized on correct cued recall. The difference is statistically significant only for pairs presented 8 times (7.28 vs. 6.10; Z = -2.93, Wilcoxon test) mainly because too few subjects correctly recall unrelated items at lower actual frequencies. These results are interesting because items show little evidence of discrimination on the basis of frequency, unless the entire study pair can be reinstated. This result suggests that frequency information does not accrue to the individual words, but to study pairs as units.

Zero items. Some discussion is necessary of the results concerning the frequency estimates given the test items that did not occur in the study list. Recall that for these items, subjects received a single word and were required to guess a word to go with it and estimate the frequency with which the pair occurred during study. Subjects were instructed to guess a "target" even if their frequency estimate was zero. Under these circumstances, the unstudied test items cannot be classed as related or unrelated. The mean frequency estimate for these distractor items was 0.90(.78). An analysis of the guesses subjects made to these unstudied test items shows that subjects produce the normatively defined "correct" related target in a substantial number

of cases. Out of 240 opportunities, 101 resulted in the subject's guessing the word that would have been the target had the pair been chosen to occur during study. Because of the way the experimental pairs were constructed, some of these guessed items appeared as unrelated partners to other items during study. Thus, a large number of the 101 "correct" guesses for unstudied test items are actually intra-list intrusions. Although this means that there is a possibility of intralist interference in these first experiments, such interference would have no bearing on the conclusions of this experiment, and would operate against the important findings of the first two experiments. No important conclusions are drawn from the results for the zero items in this experiment. In general, the fact that unrelated pairs are made of members who may have close associates in other unrelated pairs means there is an internal structure to the set of unrelated pairs that does not exist for related pairs. This internal structure could lead to a confusion about which pair members went with which. Such confusion, if it occurred, could yield reduced frequency discrimination for unrelated This lowered discriminability could mask the major difference pairs. between related and unrelated pairs obtained so far. Later experiments in this series will remove this confounding and show the same patterns of results reported so far. Finally, although subjects sometimes guess the "correct" target for unstudied test items, conditionalizing the frequency estimates for these items on "recall" performance does not yield reliable differences. Pairs for which subjects guess the "target" received mean estimates of .66 while the mean for pairs in which the "target" was not guessed was .94.

### Conclusions

It is abundantly clear that the pattern of frequency estimates is sharply altered by the experimental manipulations employed in this experiment. When frequency estimates accompany cued recall, related pairs show a consistent bias towards higher estimates. This bias parallels a marked advantage for related pairs in cued recall, suggesting that frequency estimates depend on recall. In the standard frequency estimation task subjects made frequency judgments about intact pairs, and apparently in such a situation, retrieval of pair members is not an important determinant of frequency estimates. In the more standard situation the confusion produced by pre-experimental associations between members of related pairs results in frequency estimates that are less steep but of no greater magnitude, overall. In the less conventional test situation, in which frequency estimation accompanied cued recall, the recall of pair member plays an important role in frequency estimation. Items that are more likely to be recalled show steeper frequency estimates.

Conditional analyses show that only items that can be correctly recalled show an increment in estimates with actual frequency. This result suggests that retrieval of the entire pair is crucial for successful frequency estimation because frequency information accrues to the studied pairs and not to individual members. Conditional analyses also show that correctly recalled unrelated items have a higher and steeper frequency estimation function than do related items that are scored as correct cued recall. This result suggests that some of the "correctly recalled" related targets are actually verbal associates generated at test. If this is so, the "true" frequency estimation

function for related pairs should have a steeper slope. In standard frequency estimation, conditional analyses have shown that only items that are recognized as old received accurate frequency estimates (Harris et al., 1980). When frequency estimates accompany cued recall, conditional analyses used in this experiment have shown that only items that are recalled from the study phase receive accurate frequency estimates. These conditional analyses suggest that the slope of the unconditionalized frequency estimates for related pairs is not as high as it "should" be. These results suggest that although the slopes of the cued recall functions for related and unrelated pairs do not differ, subjects have more discriminating frequency information about remembered That is, the predictions of both the availability related pairs. heuristic and the episodic/semantic hypothesis are supported. Items that are better remembered (easier to retrieve) produce higher estimates and probably better frequency discrimination.

# Experiment 4

The purpose of this experiment is to extend the findings of Experiment 2 to the test situation employed in Experiment 3. Experiment 2 established that the pattern of frequency estimates depends on the study strategy subjects employ. Thus <u>similarity</u> instructions result in higher frequency discrimination for unrelated pairs. <u>Contrast</u> instructions removed and, perhaps, reversed this difference so that related pairs showed steeper frequency estimation functions. Experiment 3 showed that the form of the frequency test. When frequency estimates accompany cued recall, related pairs receive higher frequency estimates than do unrelated pairs. This experiment will examine the two study instructions of Experiment 2 using the frequency test of Experiment 3.

According to the rationale developed so far, the associative connection between members of related pairs has an important consequence for frequency estimates. However, the consequences of the associative connection depend markedly on a number of factors. Theoretically, instructions to look for similarities cause subjects to ignore important information about the time and place of the pairs occurrence and, therefore, produce relatively undiscriminating frequency estimates. This pattern can be altered if subjects look for contrasts when they study the related pairs. A second factor is the nature of the frequency test. When frequency estimates accompany cued recall, the improved memorability conferred on the related pairs by the associative connection results in frequency estimates that depend upon recall. If the assumptions made so far are correct, the cued recall differences

produced by the similarity/contrast manipulation should be reflected in the frequency estimates that accompany cued recall. Thus any difference in cued recall performance, regardless of its source, should be reflected in frequency estimates. This expectation runs somewhat counter to the assumption (Tversky & Kahneman, 1973) that frequency estimates come from an indirect assessment of associative strength.

In Experiment 4 subjects studied a list of related and unrelated pairs under either similarity or contrast instructions. Afterwards, subjects performed a cued recall test in which they were to recall the right hand member of the study pair given the left hand member as a cue. On the same test, subjects were to indicate the study frequency of each Based on the assumptions made so far, it is expected that the pair. pattern of results for similarity processing will resemble the results That experiment showed higher recall and higher of Experiment 3. frequency estimates for related items. Contrast instructions should improve the recall of related items relative to similarity instructions, while similarity instructions should facilitate the recall of unrelated items compared to constrast instructions. Regardless of the source of the recall improvement, items that are better recalled should show higher and steeper frequency estimation relative to items that are less well recalled.

These expectations must be tempered by two considerations. First, cued recall of related items was very high in Experiment 3. It is possible that <u>contrast</u> instructions would be unable to improve on that near-perfect performance. Second, as was concluded in Experiment 3, higher frequency estimates for correctly recalled related items

than for correctly recalled unrelated items would indicate that some portion of the recalled related items were really guessed. Method

<u>Subjects</u>. Thirty-four introductory psychology students served in 2 groups of 17.

Materials and procedure. The study list and instructions were identical to those of Experiment 2. Recall that in Experiment 2, 40 related and 40 unrelated critical pairs occurred 0, 1, 2 or 3 times during study. Subjects in the similarity condition were instructed to find a characteristic pair members shared, while subjects in the contrast condition were instructed to find some way in which pair members were dissimilar. After the presentation of the list under similarity or contrast instructions, all subjects completed the same final test which was directly analogous to that of Experiment 3. Subjects were instructed to recall the right hand member of each of the 80 critical pairs (including unpresented items) given the left-hand member as a cue. Subjects were also to indicate the actual presentation frequency of the pair using the digits 0, 1, 2, or 3. The entire procedure took about 50 min.

### Results and Discussion

The results will be presented for each measure individually starting with cued recall performance followed by frequency estimates and frequency estimates conditionalized on cued recall performance.

<u>Cued recall</u>. Figure 6 shows the number of items correctly recalled. An analysis of variance upon these data yielded the typical main effect for frequency [F(2,64) = 20.29]. The analysis also yielded a main effect for instructions indicating that, overall, similarity

# Figure 6: Cued recall performance by subjects who studied under <u>similarity</u> or <u>constrast</u> instructions



processing led to better recall than contrast [mean number recalled out of 10 = 7.17 vs. 5.16; F(1,32) = 23.17, MSE = 8.89]; and a main effect of pair type indicating that related pairs led to better recall than did unrelated pairs [means = 8.77 vs. 3.55, F(1, 32) = 340, MSE = 4.10]. In addition, three interactions qualify these effects. The analysis yielded an interaction of instructions with pair type indicating that the advantage for similarity instructions over contrast only obtained for unrelated pairs; means for similarity versus contrast were 8.92 vs. 8.63 for related pairs and 5.41 vs. 1.69 for unrelated pairs [F(1,32) =The analysis also yielded interactions indicating that the 36.641. increment in number recalled with frequency was steeper for unrelated than related pairs [F(2,64) = 8.25, MSE = 1.54], and the increment in recall with frequency was steeper for similarity than for contrast processing [F(2, 64) = 3.70, MSE = .98]. These final interactions stem mainly from the fact that recall of related pairs is so high that there is not much room for improvement, and from the fact that recall of unrelated items after contrast instructions is very poor throughout. In general, study instructions had the predicted effect upon the cued recall of members of unrelated pairs. Similarity processing produced higher cued recall for unrelated pairs than did contrast processing. Probably because of a ceiling effect, related pairs were affected neither by the manipulation of study instructions nor by study frequency.

Unconditionalized estimates. The second analysis was performed on the frequency estimates collected during the final test, shown in Figure 7. There was, as usual, a main effect of actual frequency [F(2,64) = 262], and a main effect for pair type, indicating that Figure 7: Mean frequency estimates accompanying cued recall made by subjects who studied under <u>similarity</u> or <u>contrast</u> instructions



related items received higher frequency estimates than did unrelated [means = 1.84 vs. 1.53, F(1,32) = 62.43, MSE = .080]. These main effects were, again, qualified by interactions. The analysis yielded an interaction of actual frequency and pair type, indicating a steeper slope across actual frequency for related items, [F(2,64) = 4.29]. The instructional manipulation produced no main effect in this analysis but did contribute to an interaction with actual frequency indicating that the slope across actual frequency was steeper for <u>similarity</u> than for <u>contrastive</u> processing [F(2,64) = 8.49].

These results closely parallel the results for cued recall. The overall slope of the frequency estimation function was steeper for similarity processing than for contrast, and the slope was steeper for related than for unrelated pairs. Also, of course, related pairs received higher frequency estimates than did unrelated. One difference occurred in the effect of instructions. In cued recall, subjects who studied under similarity instructions recalled more unrelated items than did subjects who studied under contrast instructions. Accompanying frequency estimates for unrelated pairs, although they were steeper following similarity instructions, did not differ in overall magnitude between instructional groups. Finally, note that in Experiment 2 contrast instructions seemed to produce a bias favoring higher estimates for related pairs overall, possibly accompanied by steeper slope. Inspection of Figure 7 and the analyses reported above show that the same result was obtained in this experiment.

<u>Conditionalized estimates</u>. The final analyses concerned the frequency estimates conditionalized upon cued-recall performance (See Figure 8). The pattern does not differ with study instructions. First,

Figure 8: Mean frequency estimates conditionalized on cued recall for subjects who studied under <u>similarity</u> or <u>contrast</u> instructions



for <u>similarity</u> processing (Figure 8A), frequency estimates for pairs in which the target was correctly recalled always exceeded estimates for pairs in which the target was not recalled. For related pairs the relevant mean frequency estimates were: 1.19(.29) vs. .74(.72) for 1's (Z = -2.31, Wilcoxon Test), 2.17(.33) vs. 1.58(.64) for 2's (Z = -2.13), and 2.46(.31) vs. 1.57(1.28) for 3's (Z = -2.07). The corresponding means for unrelated items are: 1.22(.48) vs. .53(.45) for 1's (Z = -3.48), 2.06(.34) vs. 1.05(.61) for 2's (Z = -3.47), and 2.58(.26) vs. 1.53(.81) for 3's (Z = -3.42). For both related and unrelated pairs, correctly recalled items show more discriminability (a steeper slope with actual frequency) for frequency estimates than do incorrectly recalled items. In addition, this conditional analysis results in little or no significant difference between related and unrelated pairs either in discriminability or in overall magnitude.

The pattern of results is substantially the same after <u>contrast</u> study (See Figure 8B). Frequency estimates for recalled items are higher than estimates for pairs whose target was not recalled, but the differences are not always statistically significant. The means for related pairs are: 1.33(.36) vs. 1.12(.79) for 1's, 1.98(.33) vs. 1.57(.69) for 2's, and 2.33(.32) vs. 1.26(.54) for 3's, (Z = -3.04, Wilcoxon test); and the corresponding means for unrelated pairs are: 1.33(.44) vs. 1.00(.30) for 1's, 2.06(.56) vs. 1.42(.47) (Z = -2.80) for 2's, and 2.29(.59) vs. 1.78(.49) (Z = -2.24) for 3's. Part of this failure to obtain significant differences can be attributed to the fact that very few subjects recalled any unrelated items after contrastive study, and to a tendency for unrecalled items to receive higher frequency estimates after contrastive study. In general, conditionalizing frequency estimates on cued-recall performance revealed no significant differences in the estimates given related items compared to unrelated items.

Conditional analyses, thefore, confirmed expectations. Correctly recalled items showed frequency estiamtes that were higher and steeper with actual frequency than did pairs in which the target was not correctly recalled. The lack of a difference between correctly recalled related and unrelated items suggests that a relatively small number of the correct related items are in fact exclusively verbal associates generated entirely at the time of test.

# Conclusions

This experiment did yield some interesting results. First, consider only the unrelated items. Similarity study produced better recall than did contrastive study, as expected. This difference in recall performance was not paralleled by a difference in the magnitude of frequency estimates. As expected, similarity yielded a steeper slope but did not on the average produce higher estimates than contrastive study. It seems that very large differences in recall are necessary to produce differences in the magnitude of frequency estimates. Related items produced more surprising results. Cued-recall performance did not show an effect of study instructions but the accompanying frequency estimates did and in a somewhat unexpected direction. The frequency estimation function is steeper following similarity than after contrast instructions, despite the fact that contrastive study was presumed to produce better episodic memory traces for related pairs than is The recall data also did not support this assumption. similarity. Although, not significantly, subjects who studied under similarity

instructions recalled more than the subjects who studied under contrast That is, subjects who received similarity instructions instructions. produced better cued recall performance for both pair types and this may have resulted in an overall advantage in frequency discriminability for those subjects. Because cued recall of related pairs is so close to ceiling for both instructional groups, it is not possible to draw firm conclusions about the role of recall in the frequency estimates for Again, for unrelated pairs, however, the results are related pairs. Similarity instructions improve cued recall relative to clearer. This recall advantage results in a contrast instructions. discriminability and magnitude advantage for the estimates by subjects who studied under similarity instructions.

This experiment provides some information on the role of study instructions in Experiment 2. Recall that in that experiment similarity instructions produced a pattern of results close to that obtained in the That is, unrelated pairs showed better frequency first experiment. discrimination than did related pairs and there was no evidence of a bias towards the overestimation of related pairs. This pattern is taken as evidence that, in general, frequency judgments for related and unrelated pairs reflect a difference in the discrimination or sensitivity of estimates for the two pair types. Under similarity processing frequency judgments seem to show steeper slope for unrelated rather than a bias towards the relative overestimation of the frequency of related Theoretically, this is the expected result if subjects retain pairs. qualitative different information about the occurrence of each kind of study pair. Focusing on general background semantic information about related pairs should be less useful in a frequency judgment test

than the relatively specific, episodic information that would be encoded for unrelated pairs. It seems reasonable to conclude, therefore, that <u>similarity</u> processing is close to what people usually or implicitly do in such situations. <u>Similarity</u> instructions also produce the typical pattern of results when frequency estimates accompany cued recall. Recall of related targets is superior to recall of unrelated targets, and frequency estimates for related pairs are greater in magnitude and are steeper than estimates for unrelated pairs.

A change in study instructions can produce a somewhat different pattern of results. In standard frequency judgment, contrast instructions seem to produce a pattern of results most consistent with the existence of a bias toward overestimation of related pairs. There seems to be little evidence of a difference in discriminability between the estimates for the two types of pairs. In frequency estimates that accompany cued recall, the same pattern is obtained - an overall advantage for related pairs and little or no slope difference between the two estimation functions. Contrast instructions produce a study strategy that differs from the typical strategy subjects employed in Experiment 2 and the strategy subjects employed under similarity instructions in Experiment 4. Contrast instructions may cause subjects to enhance the episodic representation of related pairs, improving the quality of episodic information and permitting frequency estimation equal to (or better) than the estimates for unrelated pairs in discriminability. It is reasonable to argue that contrast instructions improve the utility of traces associated with related pairs, because, instead of focusing on general, semantic, categorical information, contrast instructions require subjects to attend to information

relatively specific to the actual occurrence of the two related words during the study phase. One may be reasonably sure that the manipulation of study instructions affected subjects' behavior because, as expected, unrelated pairs show much better cued recall after <u>similarity</u> processing than after <u>contrast</u>. In fact, recall is so poor for unrelated items after contrastive processing, and so good for related pairs overall, that the bias observed after contrastive processing may be a product of this large recall difference for the two classes of items.

Thus there is evidence for both the episodic/semantic hypotheiss and the availability heuristic (Tversky & Kahneman, 1973). Experiment 4 yielded some results that indicate the existence of a bias towards the overestimation of related pairs. Other results support the conclusion that unrelated pairs show more discriminating frequency judgments. Differences in recall probability are sometimes reflected in the magnitude of frequency estimates as predicted by the availability heuristic. Differences in recall probability are sometimes reflected in the discriminability of frequency estimates as predicted by the episodic/semantic hypothesis. There seems to be little evidence, however, that subjects base their frequency estimates directly on an assessment of the strength-of-association between pair members (Tversky & Kahneman, 1973).

### Some Review and General Discussion up to this Point

This section will provide a review of motivation for the experiments thus far and summarize the findings. First, there are a number of different patterns of results that can be obtained when subjects are asked to estimate the frequency of two classes of items

that occur a variety of numbers of times. Although the major theories of frequency estimation can predict (or at least rationalize) each pattern, given the experimental conditions that produce it, the dominant theories are unclear as to what to predict in one particular situation. What should happen if related and unrelated pairs are presented varying numbers of times each? One possibility is that the related pairs, because of the associative connection between their members, enjoy a memory advantage that would result in more discriminating (and possibly higher) frequency estimates. Another possibility is that the memory advantage for related pairs (if there is one) would really be expected to be an advantage in recall of one member of the pair given another, and that this advantage would be of little importance when intact pairs are presented for frequency estimation. This second possibility would argue that the important difference between related and unrelated pairs is that subjects have much more difficulty determining whether related pairs have occurred during study. The intimate theoretical and empirical connection between recognition and frequency estimation is already well established (Underwood, 1972; Harris et al., 1980) and this worsened old/new discrimination must result in worsened discrimination among levels of actual frequency. This means that related pairs show poorer frequency discrimination than unrelated pairs. There are a number of possible explanations for this poorer old/new discrimination for related pairs, but the one proposed by this thesis is that subjects typically encode related and unrelated pairs quite differently. Subjects typically attend to the associative connection between members of related pairs and this results in a relatively general, categorical or semantic encoding. With unrelated pairs, for which there is no

associative connection, subjects are more likely to attend to characteristics unique to the current occurrence of the pair and thereby form a relatively episodic memory encoding. If it is assumed that old/new discrimination is a necessary substage to frequency estimation, and that memory is only the trace of encoding processes, this assumption means that the encoding for unrelated pairs would have greater utility in a frequency judgment situation than the encoding of related pairs. This utility difference should result in discriminability differences in standard frequency estimation but with no important overall difference in the magnitude of the represented frequency information, especially in a within-subjects experimental design.

Experiments 1 and 2 showed that the different patterns of frequency estimates for related and unrelated pairs are not well described as a bias toward the relative overestimation of related pairs. Instead, under standard study and test conditions, at least, there appear to be no important differences in magnitude, but rather, a difference in the discrimination with which subjects can assign pairs to Unrelated pairs have a steeper frequency estimation frequencies. function, higher frequency discrimination and better sensitivity in relative frequency tests. It appears that when instructed to attend to frequency subjects engage in, more or less, the same processing strategy as they do when instructed to look for similarities or relationships between members of the study pairs. On the other hand, other study instructions produce different results. Instructions to look for contrasts or differences between members of study pairs appear to remove discriminability differences between related and unrelated pairs. After contrastive study, however, frequency judgments do seem to reflect the

existence of an overall bias favoring related pairs. This may be due to the very large recall difference that occurs under such conditions. Nevertheless, the results support the contention that under standard conditions, the important difference between related and unrelated pairs is a difference in the accuracy of the old/new discrimination necessary for frequency estimation.

Recall differences do become important in a nonstandard test. When frequency estimates accompany cued recall, recall differences become crucial. Frequency discriminability parallels differences in cued recall whether such recall differences come from differences in pair construction or from differences in study strategy. That is, it does not seem to be the associative connection, <u>per se</u>, that affects frequency judgments in this situation. Rather, differences in recall of the entire study event seem to be the crucial determinant of frequency estimates. Experiment 6 will examine this issue further.

These findings are important for any position which argues that items that are related will be judged to occur more often than items that are unrelated. This is the conclusion that was drawn by Chapman (1967, Chapman & Chapman, 1969) and others (Tversky & Kahneman, 1973; Golding & Rorer, 1972). These authors concluded that people have a bias towards overestimating the co-occurrence of things that seem to be related (or have gone together in the past, or seem like they should go together, etc.). The next experiment extends the findings for estimates of raw frequency to estimates of conjoint frequency or co-occurrence.

### Experiment 5

As mentioned above, this experiment investigates what happens if subjects estimate the frequency of co-occurrence of items rather than estimate the frequency of the pair. Chapman (1967) conducted such an experiment and concluded that subjects show a bias towards the relative overestimation of the frequency of co-occurrence of related items. As mentioned in the introduction, Chapman (1967) used only a single level of actual conjoint frequency (33-1/3%), however. For this reason, and because results of previous experiments suggest that the difference between related and unrelated items may be one of discriminability rather than overall magnitude, a re-examination of the relationship is warranted. If the important difference is one of discriminability, much of the work by Chapman and Chapman (e.g., 1969) must be questioned. It is entirely possible that the characterization of "illusory correlation" as a <u>bias</u> is in error.

In Experiment 5, every single item occurred exactly eight times. What varied was the proportion of those eight times the item occurred with another semantically related item. Conversely, the number of times items occurred with unrelated partners also varied. An item could occur with a particular related or unrelated partner on 0, 25, 50, 75, or 100% of its occurrences. After study, subjects were presented with a list of pairs and were required to estimate what proportion of occurrences of the left-hand member were accompanied by an occurrence of the right-hand member. In other words, "Given the left-hand member, what was the probability that the right-hand member would also occur?".

The predictions for this experiment follow directly from previous considerations. Because related pairings are encoded in terms of general, categorical or semantic characteristics and unrelated pairings are encoded in terms of features unique to the event's occurrence, subjects should have better memory information (for the purpose of old/new discrimination and, consequently, frequency judgments) about unrelated pairings. Conjoint frequency estimates for unrelated test pairings should show greater discriminability across levels of actual frequency, and a steeper estimation function. Because this experiment used a within-subjects design, it was expected that subjects would retain a fairly good idea of the relative number of related and unrelated pairings during study, and, therefore, no important differences in the overall magnitude of conjoint frequency estimates were predicted. These predictions have an important implication. If unrelated pairings show better frequency discrimination than related pairings, and if there are no overall differences in magnitude, related pairings must receive higher estimates than unrelated pairings at relatively low levels of actual conjoint frequency. As mentioned in the introduction, many experiments demonstrating illusory correlation employ zero or low actual values. Clearly, if only low levels of actual frequency and correlation are examined, discriminability differences could look like an overall bias. The theoretical position advocated here predicts a discriminability advantage for unrelated pairings over related pairings, as that if only a single low actual conjoint frequency is examined, results consistent with the bias view will be obtained.

This experiment also employed a manipulation of study instructions. Some subjects were told to try to remember the frequency of co-occurrence for the various study items and they were told they would receive a test of conjoint frequency afterwards. The other subjects were told only that study items would sometimes appear with different partners and that there would be a later memory test. There is considerable evidence that this kind of study manipulation has little effect on frequency estimates (Harris et al., 1980; Howell, 1973) but this experiment examined the effects on estimates of conjoint frequency.

Basically, the design consisted of a within-subject manipulation of pairing type (related or unrelated), a within-subject manipulation of actual conjoint frequency (0,25,50,75,100%) and a between-subjects manipulation of study instructions (conjoint frequency or general memory). There were, in fact, two different types of pairings within the 25% level, but the main interest lies in only one of those. <u>Method</u>

<u>Subjects</u>. Forty-eight introductory psychology students from McMaster University served as subjects in partial fulfillment of a course requirement. There were two sets of study instructions, each given to 24 subjects.

Materials and procedure. Sixty pairs were chosen from Palermo and Jenkins (1964) such that the second member of each pair was the highest verbal associate of the first. The 60 pairs were randomly divided into 6 groups of 10. Every single word in the experiment occurred exactly 8 times. For the first group of 10, the intact related pairs were repeated 8 times each in the study list. For the second group, the pair members were randomly reassigned to new partners, and

these unrelated pairs were repeated intact. For the third group, the left-hand members were presented 4 times with their related partner and 4 times with an unrelated partner (the same partner each time). For the fourth group, the left-hand items were presented twice with their related partners and twice each with 3 other unrelated partners. For the fifth group, the left-hand members were presented 6 times with their related partner and twice with an unrelated partner. For the sixth group, the left-hand members were presented twice with their related partners and 6 times with an unrelated partner. The presentations of left-hand members (whether with related or unrelated partners) were randomly ordered throughout the study list. Mean spacing between repetitions of left-hand members was 60 intervening items. Groups of filler pairs occupied the 10 first and last positions in the long study list. The entire 500-item list was videotaped at a rate of 4.5 s/item.

Subjects watched the study list under one of two sets of instructions. Half of the subjects were told they would see a long list of word pairs in which words would be repeated from time to time. The words could appear in the company of the same partner or with different partners. Subjects were told that the pairs would occur at a fast rate and were told they would have to concentrate because there would be a final test that would require them to say, given two words, how often they occurred <u>together</u> in the list. The other 24 subjects received the same instructions except that they were told to expect a final test of memory for the words in the list.

All subjects received the same final test consisting of 120 questions. Each question consisted of two words separated by a blank. Subjects were instructed to write in the blank (expressed as a

percentage) the proportion of times that, given the first member occurred in the study list, it was followed by the second. Every lefthand member from the experiment was tested twice, once with its related partner and once with an unrelated partner. The test was 2 pages long and each left-hand member occurred once on each page. Because subjects were instructed to work fairly quickly and give their first impression as to the proportion of co-occurrences for each pair, the entire procedure required about 50 min.

### Results and Discussion

The major analysis for this experiment concerned the mean estimates of conjoint frequency (expressed as percentages). These results are summarized in Figure 9. Note that these results closely resemble frequency estimates obtained in experiments involving absolute estimates of intact pairs (e.g., Figure 2, similarity instructions). There were 5 levels of actual conjoint frequency (0, 25, 50, 75, and 100%). This analysis ignored the 25% level in which the other member of the pair varied (group 4 of the method); a comparison of the two 25% manipulations is considered later. The mean conjoint frequency estimates for subjects who received general memory instructions were subjected to an analysis of variance in which pair type (related or unrelated) and actual conjoint frequency were repeated measures. The analysis yielded a significant main effect of actual conjoint frequency [F(4,92) = 157]. Also, there was an effect for item type indicating that, overall, related pairs received higher estimates [means = 45.6% vs. 42.1%; F(1,23) = 4.29, MSE = 166]. The analysis also yielded an interaction of these two factors, indicating that the advantage for related p[airs was reversed at the higher levels of actual conjoint
Figure 9: Mean conjoint frequency estimates for subjects who studied under A, general memory; or B, specific correlation

instructions



ACTUAL CONJOINT FREQUENCY (%).

frequency, and that the slope of the function relating estimates to actual values is steeper for estimates made for unrelated pairs [F(4,92)= 2.66, MSE = 104]. An identical analysis was performed on the estimates made by the subjects who received the correlation instructions with the identical result. Again there were significant main effects of actual conjoint frequency [F(4,92) = 147] and of item type; related pairs received higher estimates than unrelated [means = 47.3% vs. 43.5%, F(1,23) = 5.32, MSE = 161]. Again, this analysis yielded the same crossover interaction of actual frequency and item type [F(4,92) = 4.10, MSE = 110]. There was no apparent difference attributable to the manipulation of study instructions.

The next analysis concerned the two conditions in which the actual conjoint frequency was 25%. In one case (consistent) the tested items were studied together twice and with the same other item six In the other case (inconsistent) the items occurred with three times. others, twice each during study. When tested with a related word, six of the item's other occurrences had been with unrelated words. When tested with an unrelated word, all other occurrences had been with a single related word in the consistent condition, and in the inconsistent condition other occurrences had been with a related word twice and with two other unrelated words twice each. The mean conjoint frequency estimates (out of 100%) were subjected to an analysis of variance in which study instructions, item type (related or unrelated) and presentation format (consistent or inconsistent) were all withinsubjects factors. As would be expected from the previous results, this analysis yielded a significant effect for item type such that related pairings received higher estimates of conjoint frequency than did

unrelated [means = 34.5 vs. 27.3, F(1,46) = 18.85, MSE = 131]. This effect was qualified by an interaction with presentation format such that the advantage for related pairings was greater when all pairings during study had been 25%. The means for the 75% condition were 31.7 vs. 28.6, and 37.3 vs. 26.1 for the 25% condition [F(1,46) = 16.54, MSE = 46.4].

Thus, if only one level of actual conjoint frequency is examined, results are obtained that make it appear as if there is a bias toward the overestimation of related pairs. Related pairs received higher estimates than unrelated and they received estimates that were much higher than the actual conjoint frequency (always 25%). In this sense, unrelated items actually received more accurate estimates. At low levels of actual frequency, more accurate estimates for unrelated pairs can appear to be a general bias toward the overestimation of related pairs. Only by examining a wider range of actual frequencies is it possible to discriminate between the two alternatives. Finally, study context can have some effect on conjoint frequency estimates, but this effect does not alter the basic finding. Consistent study pairings (either related or unrelated) reduce the size of the advantage for related pairs at low levels of actual frequency. It seems probable, however, that inconsistent pairing is a closer approximation to "realworld" conditions.

## Conclusions

The expectations for this experiment were confirmed. First, both study groups yielded identical patterns of results and the two groups can be considered replications of each other. The estimates of conjoint frequency yielded results exactly comparable to those obtained

with judgments of absolute frequency. Although there was a small effect of item type such that, overall, related pairs received slightly higher estimates than unrelated, the most important finding is that of an interaction of item type with actual conjoint frequency. This interaction indicates that subjects are able to give more accurate estimates to unrelated pairings. This is the result one would expect if there was any confusion between intra-experimental and pre-experimental Theoretically, when subjects co-occurrences of the studied items. encounter a related pair they do different things than when they encounter an unrelated pair. Processing or encoding a pair of related items in terms of well-known stereotypic characteristics results in a memory trace that is relatively semantic in content. The processing of unrelated items is different. There is no easy way to process unrelated items in a stereotypic fashion, hence the resulting trace tends to be unique to the time and place of the event's occurrence. Episodic information is of greater utility in performing a subsequent test of frequency which is, by definition, episodic in nature. Subjects retain a fairly good representation of the average level of frequency for the two classes of items so that the important difference between related and unrelated pairings is that the latter receive frequency judgments with greater discriminability across actual frequency and usually greater accuracy.

The observation that the difference in estimates of conjoint frequency between related and unrelated pairings is not mainly a bias for related but an increase in accuracy for unrelated pairings contradicts the conclusions of Chapman and Chapman (1969) and Golding and Rorer (1972). These experiments typically expose subjects to low, zero or even negative correlations between ostensibly related items. If there is error in the judgment process, any estimation of frequency (or correlation) for low actual levels <u>must</u> result in overestimation - the greater the error the greater the mean overestimation - at low levels only. Thus, if a range of actual conjoint frequencies is not explored, a difference in the accuracy of estimates would masquerade as an overall bias. The issue may not be crucial for clinical situations (Chapman & Chapman, 1969; Golding & Rorer, 1972) where overestimating non-existent correlations may be more serious than underestimating very high correlations. The issue <u>is</u> crucial, however, for an understanding of the representation and use of frequency information. It is entirely possible, therefore, that the phenomenon of illusory correlation is actually a difference in judgment sensitivity and not a bias.

## Experiment 6

This experiment was conducted to investigate the direct role of the associative connection between pair members on frequency judgments. In their discussion of the availability heuristic, Tversky and Kahneman (1973) conclude that people make frequency estimates based on ease of retrieval (availability) and on the "strength of the associative connection" between two events. Earlier experiments have shown that, under some circumstances, subjects do seem to base frequency estimates upon recall. It is also undeniable that ease of retrieval will depend strongly upon associative connection in most situations. But, unless one is willing to assert that subjects have precise information about the relationship between retrieval and associative strength, postulating that frequency estimates are inferred from the strength of association is postulating a mechanism different from availability. Experiment 4 provided evidence that the discriminability of frequency estimates sometimes depends upon availability, whether availability stems from semantic associations (associative connections) or from episodic manipulations of study instructions. This experiment will show that differences in associative connections do not necessarily produce differences in frequency estimates.

An experiment by Mathews (1977) provided a means to accomplish such a manipulation of the relationship between members of pairs. Mathews presented pairs of words like <u>lion-whale</u> and <u>elephant-trapeze</u> along with encoding categories like PART OF A CIRCUS or A MAMMAL. The subjects' task was to indicate whether both, one or neither member of the word pair were members of the category. Mathews constructed

counterbalanced lists in which all the same pairs occurred, but in each list the number of pair members relevant to the encoding category varied. After such a study list, Mathews obtained free and cued recall. Mathews' results showed that recall success depended directly on the degree of relation. Pairs in which both items were members of the encoding category produced better recall than pairs in which one item was a category member, and pairs in which neither item was a category member produced the poorest recall of all. The particular semantic attributes focused on during study greatly affected later usefulness in recall. Mathews also concluded that recall was accomplished via the mediation of the encoding category, even for the categories in which neither study item was a member.

If frequency estimates are inferred directly from the degree of semantic association, those study pairs that are studied so that both words are exemplars of the encoding category should receive higher frequency estimates than pairs in which only one word is a member of the category. On the other hand, if the strength of associative connection, <u>per se</u>, is not the basis for frequency estimation, no overall magnitude differences are necessarily predicted. In addition, if slope differences observed in earlier experiments are due to different encoding strategies and the different utilities of the episodic or semantic memory information encoded under such strategies, then no discriminability differences would be expected in such an experiment either. In order to answer the orienting question for each study pair, the subject must always attend to the specific encoding category presented with each pair. This requirement means that the subject cannot deal with any pair in the obvious semantic fashion that is,

theoretically, the usual encoding for related pairs. Because all study pairs require the same attention to the actual presentation episode, no slope, discriminability or accuracy differences are expected between pairs in which both words are category members and pairs in which only one member is a category exemplar.

Some additional assessment of the role of recall in frequency estimation might be provided by frequency estimates for items in which neither member of the study pair is an exemplar of the study category. Mathews (1977) concluded that recall of study items was mediated by recall of the study category. If this is so, recall of the study category would be considerably worse for pairs that have no members in the study category than for other pairs. That is, a test pair in which one or both of the members is a category exemplar should have a much higher chance of producing recall of that category than a pair in which neither member is a category exemplar. If this is so, and if frequency estimates depend upon recall of the entire study episode (items + category), then frequency estimates for these zero pairs should be lower than the estimates for the other two classes of pairs. A finding that the zero pairs do receive lower and less discriminating frequency estimates would be similar to a more general finding, namely that items studied under orienting questions that receive negative answers often show poorer retention than items that receive affirmative orienting answers (Craik & Tulving, 1975). An explanation of such a result in terms of differences in the recall of the orienting task is also consistent with the explanation in terms of elaboration differences in the original encoding (Craik & Tulving, 1975).

Experiment 6 also included frequency estimation for single study items. Again, if frequency estimates depend upon retrieval of the entire study episode, then single words from pairs in which both members were category exemplars should have the highest probability of producing retrieval of the study category and the other pair member. The retrieval probability for single items from pairs in which only one member was a category exemplar should be lower, because half of these items are not members of the encoding category. The retrieval probability for single words from study pairs in which neither member was a category exemplar should be lower still. If, as suggested above, frequency estimates depend upon retrieval of the entire study event, frequency estimates for single words should mirror the presumed retrieval of entire pairs.

Subjects in this experiment studied pairs of words under conditions in which both members, or one member, or neither member of the study pair was a member of some orienting semantic category. During study, each subject's task was to report the number of pair members that were members of the category. Three counterbalanced lists were used to ensure that each pair was studied under each appropriate orienting condition. Following study, subjects estimated the frequency of all experimental pairs or single words. A view which states that frequency estimates are directly inferred from the degree of semantic association predicts higher estimates for pairs in which both members are exemplars of the orienting category.

Method

<u>Subjects</u>. Seventy-two McMaster introductory psychology students participated in six groups of 12 in partial fulfillment of a course requirement.

Battig and Montague (1969) and other sources were Materials. culled in order to construct 27 word groups such that the following relationship held: each group consisted of two pairs of words and a pair of labeled characteristics (see Mathews, 1977); when presented with one label, both members of one pair and one member of the other pair possessed the characteristic; when presented with the other label, one member of the first pair and both members of the second pair possessed the second characteristic. Finally, for each pair of words, a label from a different word group was chosen so that neither member possessed the characteristic. The example in Table 6 shows the structure of the materials. One group of subjects studied the pair Mixmaster-Power saw with the orienting label ELECTRICAL. Both members of this pair are electrical devices and, therefore, share the orienting Another group of subjects studied the same pair, characteristic. Mixmaster-Power saw, with the orienting label IN A KITCHEN. In this case, only one pair member possesses the orienting characteristic. Finally a third group studied Mixmaster-Power saw with the label MAMMAL, characteristic neither pair member possesses. This structure permitted each pair to be presented to each subject and allowed the variation of the orienting characteristic, and, consequently, the number of pair members that possessed the orienting characteristic. the structure of the lists also ensured that each subject studied each orienting label. Subjects who received Mixmaster-Power saw with the

Table 6: Examples of triplets in the three encoding conditions of Experiment 6 (After Mathews, 1977)

Number of pair members that are members of the encoding category<sup>1</sup>

1

0

Mixmaster Power Saw

ELECTRICAL IN A KITCHEN MAMMAL

Mixmaster Power Saw .

Mixmaster Power Saw

2

IN A KITCHENELECTRICALIN A FORESTSpoon StoveSpoon StoveSpoon Stove

<sup>1</sup>Orienting categories are shown in upper case. Any single subject saw a given pair in only one encoding condition.

label ELECTRICAL also studied <u>Spoon-Stove</u> with the characteristic IN A KITCHEN. Meanwhile, subjects who studied <u>Mixmaster-Power saw</u> with the label IN A KITCHEN, received <u>Spoon-Stove</u> with the characteristic ELECTRICAL. This structure ensured that all subjects studied each pair and each orienting label an equal number of times but permitted the orthogonal variation of the number of pair members that possessed the orienting characteristic. A set of 18 unpresented pairs and 20 filler pairs were also constructed by similar means. Of the 54 pairs of words, 18 were randomly chosen to occur once in the study list, 18 to occur twice and 18 three times.

The counterbalanced lists were constructed such that each pair occurred in the identical position on each list. The only difference between lists was the nature of the labeled category associated with each pair. On each list, a different third of the pairs at each frequency was always associated with the label corresponding to the shared characteristic, a different third was always associated with the label associated with the characteristic possessed by only one member, and a different third with the characteristic possessed by neither member.

The counterbalanced lists were videotaped at a rate of 10 s/pair. For each pair, the appropriate labeled characteristic was presented aloud by the experimenter. Filler items comprised approximately the first and last 25 positions of the 171 item list.

<u>Procedure</u>. Each of the three counterbalanced study lists was presented to 24 subjects. In every case, during study, the subject's task was to indicate, on a separate response sheet, how many members of each pair (0, 1 or 2) possessed the characteristic mentioned.

After study, half of the subjects receiving each study list completed a final pair frequency estimation test. The 54 critical pairs plus 18 unpresented pairs were printed on a sheet. The subject's task was to indicate the frequency (0, 1, 2 or 3) with which the pair had occurred during the first phase of the experiment. The other 12 subjects from each group completed a final single item frequency test in which the subject's task was to give a frequency estimate (0, 1, 2 or 3) for each of the 108 critical single words and 36 unpresented items used in the experiment. The entire procedure required about 45 min.

## Results

Before analyses were conducted, the response sheets for the study phase were examined. Only subjects who disagreed with the experimenter-defined answers on fewer than 10% of the responses were included in the analyses. The mean error rates for the pair test group was 4.8%, and was 4.5% for the single word test group. This screening resulted in the removal of 9 subjects from the analyses, and 3 others were randomly rejected to permit analyses on groups of equal size.

<u>Single words</u>. The first analysis concerned the frequency estimates for single words. These data are summarized in Figure 10 which shows that items from pairs in which one or both members possessed the study characteristic produced frequency estimates that were higher and steeper than items from pairs in which neither member possessed the study characteristic. The mean frequency estimates for the 108 critical words (excluding distractors) were subjected to an analysis of variance in which actual frequency and the number of characteristics present in study (0, 1, 2) were within-subjects factors. The analysis yielded a main effect for frequency [F(2,58) = 346], and a main effect for the

Figure 10:

Mean frequency estimates for single pair members that came from study pairs on which both members, one member, or neither member had been instances of the orienting category



number of characteristics shared by members of the pair from which the word came. When neither of the pair members possessed the characteristic, frequency estimates were lower (mean = 1.41) than when the word came from a pair in which either one or both members possessed the relevant characteristic [means = 1.79 and 1.81, respectively, F(2,58) = 36.98, MSE = .123]. The two main effects also contributed to an interaction such that the frequency estimates for words, from study pairs in which one or both of the members possessed the characteristic, show a steeper slope across actual frequency than do estimates for words from pairs not possessing the characteristic [F(4,116) = 3.17, MSE = .062].

<u>Pairs</u>. The second analysis concerns the frequency estimates given the 54 critical pairs for the experiment. These data are summarized in Figure 11. Again, mean frequency estimates were treated by an analysis of variance. The analysis yielded a main effect of actual frequency [F(2,58) = 758], and a main effect for item type, such that pairs in which one or both of the members possessed the relevant characteristic during study received higher estimates [means = 1.98 and 1.95, respectively] than did pairs in which neither member possessed the study characteristic [mean = 1.72, F(2,58) = 17.13, MSE = .111]. The interaction of these two factors was not significant.

### Conclusions

The conclusions for this experiment are simple. It is not the existence of a semantic relationship <u>per se</u> that produces the typical result of slope differences between related and unrelated pairs. In this experiment, the degree of semantic relation for pair members was manipulated by altering the orienting category. This manipulation of

Figure 11:

.

Mean frequency estimates for pairs in which both members, one member, or neither member had been instances of the orienting category



semantic relationship produced no accuracy differences between pairs in which either one or both members possessed the orienting characteristic. No slope differences were detected because in both conditions, subjects engaged in the same processing task. The decision that one or both members possess the study characteristic results in equivalent traces with respect to the episodic/semantic content. This equivalence of episodic/semantic information yields frequency estimates of equivalent magnitude and accuracy. One might expect the decision that no members possess the orienting characteristic to be somewhat more easily accomplished on the basis of general semantic information (given the way orienting categories were assigned to pairs). In fact, although there was not a significant slope difference, pairs in which neither member possess the orienting characteristic showed estimates of lower magnitude and consequently generally lower accuracy than the other two classes of test pairs.

The results for single items also tend to confirm this position. Apparently the accrual and use of frequency information is equivalent for items from pairs in which one or both members possessed the study characteristic because both classes of items have undergone equivalent processing in terms of use of episodic and semantic information. Frequency estimates for items from study pairs in which neither member possessed the orienting characteristic were lower in magnitude and in frequency discrimination (had a lower slope) than estimates for the other two classes if items. This difference may be due to the fact that these items from zero pairs are actually less well remembered. This memory difference could stem from two sources. First, Mathews (1977) argued that the orienting category label mediated recall of pair members. It is quite conceivable that recall of the orienting label would be poorer for zero pairs and, because frequency information accrues to whole pairs, frequency estimates for zero items would show the pattern obtained here. Second, because the processing decision for the zero pairs is, in a sense, more trivial than for the other classes of pairs, memory for those items would be worse. For whatever reason, decision difficulty seems directly related to memory (Jacoby, Craik & Begg, 1979), and it seems that frequency estimates for single items may not follow exactly the pattern of results obtained for intact pairs. The next experiment will investigate this possibility further.

Manipulation of the degree of association had little direct effect on frequency estimates. It certainly did not produce the characteristic related/unrelated accuracy differences or results that look like there is a bias towards overestimating related pairs. It could not be argued, from the data of this experiment alone, that frequency estimates are directly inferred from the degree of semantic relation between pair members. The pattern of results obtained here is decidedly not the pattern obtained in standard situations. Therefore. any assertion that the accuracy differences in standard situations results from the direct influence of the semantic relationship can be discounted. This conclusion does not deny the availability hypothesis in its basic form. Subjects do base frequency judgments upon retrieval success when there is something to retrieve. There is no need to assume, however, that frequency judgments are also directly based on correlates of recall such as associative strength or semantic relatedness.

# Experiment 7

This final experiment was conducted for several reasons. The first purpose is an examination of the accuracy of frequency estimates. As mentioned in the Introduction, accurate frequency estimates must be of the appropriate overall magnitude as well as sufficiently Subjects in all the experiments reported thus far discriminating. tended to underestimate actual frequency when actual frequency was high, while estimates for lower actual frequencies tended to be very accurate. The typical frequency estimation function tended to be negatively accelerated rather than a straight line. Because estimates for high frequency items were so low, the fact that, at high actual frequencies, estimates for unrelated pairs were higher than for related, made it seem that estimates for unrelated pairs were more accurate. It may be, however, that estimates for unrelated pairs would be higher even if subjects overestimated at high actual frequency. If this happened, estimates for unrelated pairs would actually be less accurate. In order to separate frequency discrimination (slope of the estimation function) from accuracy (closeness to actual values), subjects in Experiment 7 were permitted to overestimate actual frequency at all levels. Subjects could give estimates as high as 6 but any pair could occur at most only 4 times.

Second, two relatively new measures of frequency were included. First, in order to assess the time course of the discriminability differences, continuous on-line frequency estimates for related and unrelated pairs were collected. Such on-line estimates are typically extremely accurate (Begg, 1974), but if the slope and discriminability

differences obtained in earlier experiments are due to differential efficiency of encoded information, trends toward slope difference should appear even during continuous estimation. The other measure was frequency estimates for single pair members. The results of Experiment 6 suggested that estimates for individual words were mediated by retrieval of the study event. If so, estimates for single words in Experiment 7 should reflect the probability of pair recall given one member as a cue. That is, unlike estimates for pairs, estimates for single items should result in higher estimates, overall for words that had been members of related study pairs.

Third, as a mentioned before, there is a possible problem with the experimental materials used in the first few experiments. The unrelated study pairs were constructed by randomly re-pairing members of related pairs. This means that, although the members of unrelated pairs were not related to each other, each could be related to one member of some other unrelated pair. This kind of inter-pair structure is forced by the correlational design of Experiment 5 and is probably inevitable in the design of Experiment 6, but is certainly not necessary in the use of intact study pairs. The inter-pair relations among unrelated study pairs may have introduced a degree of intra-list interference for unrelated pairs. By the episodic/semantic encoding theory advocated in this thesis, intralist interference should mitigate against producing the discriminability advantage predicted for unrelated pairs. Although results consistent with this discriminability advantage were obtained in the early experiments, the effects were sometimes not significant. In Experiment 7, the experimental materials were constructed to avoid inter-pair relations in order to remove a source of interference which

may have attenuated the expected results from Experiments 1 through 4. In order to ensure that the conclusions of the first four experiments are not qualified by the change in the construction of study pairs, all the measures of frequency judgment were included in this experiment.

Fourth and finally, Experiment 7 involved a manipulation of study instructions. As mentioned above, some subjects engaged in online frequency estimation. It was concluded after Experiments 1 and 2 that subjects instructed to attend to frequency during study nevertheless engaged in an encoding strategy that involved some kind of attention to associative connections between pair members. Experiment 7 actually compared subjects instructed to attend to semantic relationships to those instructed to attend to frequency. It is a fundamental prediction of the episodic/semantic encoding account that both instructions produce the same pattern of results in all the frequency judgment measures. What should happen if subjects attend, not to running frequency or associative connections, but to the frequency with which pair members occur together in everyday life? Subjects should notice the relation between words like lion-tiger and rate them as occurring together more often than words like apple-shoe. What are the consequences of this decision, however? In some sense, at least, subjects are equally required to attend to semantic, linguistic characteristics and to episodic characteristics for a background frequency decision for both related and unrelated pairs. That is, episodic/semantic encoding view can predict that no differences in discriminability would occur with subjects who study under this third instruction.

Accordingly, Experiment 7 included a study list composed of related and unrelated pairs presented 1, 2, or 4 times each, and a manipulation of study instructins. Some subjects studied the list by estimating frequency on line. It is expected that these continuous estimates will yield more discriminating estimates for unrelated than for related pairs. Other subjects were required to judge the degree of association for each study pair. It was expected that the pattern of final test frequency estimates would be the same as that of the final test from the running frequency subjects. A third instructional group estimated the degree to which the members of the study pairs occur together in everyday experience. It was expected that these subjects would produce frequency estimates for related pairs that did not differ in accuracy from estimates for unrelated pairs. Some subjects completed a final test of frequency for single items with the expectation that, because those words that had been members of related pairs would have a higher probability of reinstating the entire study pair, single items from related pairs would receive higher frequency estimates overall. The pattern is expected regardless of study instructions. Other subjects were tested with cued recall accompanied by frequency estimation. Again the expectation was that related pairs would show higher (and possibly more accurate) frequency estimates than unrelated pairs regardless of study instructions. Finally, still other subjects performed a final test of relative frequency like those in Experiments l and 2. In this test subjects were presented with 2 pairs - both related, both unrelated or mixed - in which either pair could have occurred any number of times (including zero) during study. The subject's task was to indicate which pair occurred more often in the

study list. The foregoing analysis for study instructions and absolute estimates of pair frequency provided the expectations for this measure. Because subjects have less accurate information about related pairs, relative frequency comparisons involving related pairs should produce more errors than comparisons involving unrelated pairs. This pattern should not be obtained for subjects who study under background frequency instructions.

In sum, the design of Experiment 7 involved a between-subjects manipulation of the form of the final test (relative frequency estimates for pairs, absolute estimates for pairs and single words, estimates accompanying cued recall), a between-subjects manipulation of study instructions (running frequency, background frequency, degree-ofassociation), a within-subjects manipulation of study pair type (related, unrelated), and a within-subjects manipulation of actual study frequency (0, 1, 2, 4).

#### Method

<u>Subjects</u>. One hundred and eighty McMaster introductory psychology students participated in partial fulfillment of a course requirement.

Materials. From Palermo and Jenkins (1964), 114 pairs of words were chosen such that the second member was the highest verbal associate of the first member. The pairs were divided into 72 critical pairs and 28 fillers. The 72 critical pairs were further divided into 24 intact related pairs and 48 remainders from which 24 unrelated pairs were constructed by assigning, at random, half of the first members to the second members of the other half of the set. From each group of 24 pairs, 6 were chosen not to occur in the study list, 6 to occur once, 6

to occur twice, and 6 to occur 4 times. Twenty-eight pairs of filler items were constructed in the same way; these fillers occupied approximately the first and last 20 positions in the 156 item study list. The presentation list was videotaped at a rate of 8 s/pair, with an average spacing of 14 intervening pairs between repetitions.

<u>Procedure</u>. Subjects studied the list under one of 3 sets of instructions. One third of the subjects were instructed to perform an on-line frequency estimation task, judging the current presentation frequency of each pair as it occurred in the list. Another third of the subjects rated each pair, as it occurred, on "the degree with which the two members occur together in everyday experience." Rating was on a 6 point scale where 6 corresponded to "(almost) always" and 1 to "(almost) never." The final group of 60 subjects rated each pair on a scale of 1-6 on the "degree to which the members were associated or related in meaning." A rating of 6 corresponded to "very high association" and 1 to "no association."

For each group of 60 subjects, the distribution of final tests was identical. For each group, 30 subjects completed the relative frequency discrimination test. There were 3 counterbalanced forms of this test, each completed by 10 subjects in each instructional group. Each form presented sets of two pairs of critical items (including distractors), such that all possible pairings of actual frequency and item type (related/unrelated) appeared. The subjects' task was to circle the pair that had occurred more often in the study list.

For each instruction, another group of 10 subjects performed the final cued-recall test. The left-hand members of all 48 critical pairs were presented and the subject's task was to recall right-hand members

and indicate the frequency with which the pair occurred in the study list (0-6).

The final group of 20 subjects for each instruction completed a conventional frequency estimation test. Ten received a sheet consisting of the 48 critical pairs, and 10 received a sheet consisting of the 96 (unpaired) critical single words. In each case the subject's task was to indicate the frequency (0-6) with which the item had occurred in the presentation list.

In all, the experiment consisted of a between-subjects manipulation of on-line study task, and a between-subjects manipulation of final test type, with within-subjects manipulations of item type and actual frequency. The entire procedure required about 40 min.

## Results and Discussion

Because of the large number of measures involved, results will be reported for each measure separately. First, on-line frequency estimates are discussed, followed by estimates for single words, estimates for pairs, relative frequency and the various measures associated with the cued recall task. For each measure, comparisons between instructional groups will also be discussed while comparisons between measures will be reserved for the end of the section.

<u>On-line estimates</u>. The first results of interest concern the on-line frequency estimates made by subjects in the running frequency study condition. These data are summarized in Figure 12, which shows that, despite very accurate performance, unrelated pairs received more accurate estimates than did related pairs. At every level of actual frequency, unrelated pairs had mean estimates closer to the true value than did related pairs. These conclusions are supported by an analysis <sup>3</sup> Figure 12: Mean on-line frequency estimates for 30

subjects



of variance conducted on mean frequency estimates for 30 subjects. Actual frequency and item type (related or unrelated) were withinsubjects factors. As usual, the analysis yielded a main effect for actual frequency [F(3,87) = 1229]. Actual frequency interacted with item type, with the result that unrelated pairs received higher frequency estimates at the highest level of actual frequency [F(3,87) =59.20, MSE = 0.25]. Because the largest difference between related and unrelated pairs occurred at the highest level of actual frequency, unrelated items received higher estimates than related items overall, [means = 2.56 vs. 2.37, F(1,29) = 28.06, MSE = .075]. Despite highly accurate performance and very low variability, unrelated items show better discrimination on the basis of frequency than do related pairs.

Absolute estimates - singles. Mean estimates for single words for the subjects in all three instructional groups are shown in Figure 13. Although there are differences in overall magnitude of estimates, the same pattern of results is obtained for each group. As expected, items from related pairs received higher estimates overall, and these estimates for related items showed sharper discrimination (steeper slope) across actual frequency than did items from unrelated pairs. These conclusions are supported by an analysis of variance conducted on the mean frequency estimates for each subject. In this analysis, study instruction was a between-subjects factor and actual frequency and item type were within-subjects factors. The analysis yielded a main effect for instructions [F(2,27) = 6.9], MSE = 1.29], but this effect contributed to no interactions. The analysis yielded a main effect for actual frequency [F(2,54) = 198, MSE = .364] and a main effect for item type, indicating that items from related study pairs received higher

Figure 13:

Mean frequency estimates for individual pair members made by subjects who had studied by assessing A, the semantic relationship; B, background pair frequency; or C, running frequency



estimates overall than items from unrelated pairs [means = 1.68 vs. 1.34, F(1,27) = 38.47, MSE = .275]. These two main effects contributed to an interaction indicating a steeper slope across actual frequency for items from related study pairs [F(2,54) = 4.98, MSE = .141].

Absolute estimates - pairs. The pattern for final frequency estimates of intact pairs, shown in Figure 14, is more complicated. Except for differences in overall magnitude, the study condition in which subjects assess the degree of association and the condition in which subjects estimate running frequency produce the same pattern of results: Instead of differences in the overall magnitude of estimates for related and unrelated pairs, the two item classes produce frequency estimates that differ in the steepness of the function relating estimates to actual frequency. Frequency estimates for unrelated pairs are more accurate than estimates for related pairs. The pattern for the remaining study condition is slightly different. When subjects studied the pairs by assessing background frequency, final estimates for related pairs appear to be higher than for unrelated pairs. There appears to be no real slope difference between the two functions, but the mean estimates for unrelated pairs are closer to actual values than are the estimates for related pairs. That is, estimates for unrelated pairs appear to be more accurate than those for related pairs. These conclusions were supported by an analysis of variance conducted on the mean frequency estimate for each subject. Actual frequency and pair type (related or unrelated) were withinsubjects factors and study instructions was a between-subjects factor in the analysis. The analysis yielded a main effect for instructions,

Figure 14:

Mean frequency estimates for pairs made by subjects who had studied by assessing A, the semantic relationship; B, background pair frequency; or C, running frequency


indicating differences in overall magnitude of estimates between the groups [F(2,27) = 10.96, MSE = .714], and a main effect of actual frequency [F(2,54) = 461, MSE = .241]. There was no main effect of pair type, but this factor interacted with actual frequency, indicating a steeper slope overall for frequency estimates of unrelated pairs [F(2,54) = 28.12, MSE - .119]. Actual frequency also interacted with instructions, indicating that subjects who studied by assessing running frequency produced estimates that showed less discrimination across actual frequency than did the other subjects [F(4,54) = 3.20, MSE =.241]. Finally, the analysis yielded a significant 3-way interaction indicating that the slope difference between related and unrelated pairs was not obtained in the estimates of subjects that studied by assessing background frequency [F(4,54) = 3.04]. A post-hoc analysis was conducted to compare the frequency estimates for related and unrelated pairs for this background - frequency group only. This post-hoc analysis indicated no significant difference between the two conditions. Thus it is unclear whether the background-frequency study produced a bias favoring related pairs. The functions for related and unrelated pairs for this group appear to converge at the highest level of actual frequency and if a wider range of actual frequencies were sampled the typical interaction might be obtained. Nevertheless, it does seem clear that there are no important slope or discriminability differences between the frequency estimates for related and unrelated pairs after subjects study the pairs by assessing background frequency.

<u>Frequency discrimination</u>. The next results of interest concern the frequency discrimination measures. First of all, consider those comparisons in which actual frequency was equal and the two pairs Table 7: Proportion of related choices for pairs of equal actual frequency.

ACTUAL F	REQUENCY
----------	----------

		0	. 1	2	3
	Degree of Ass'n	•65*	.52	.47	.37*
STUDY CONDITION	Background F	.52	.67*	.48	.40
	Running F	.53	.57	•27*	•23*

\*proportion significantly from .5 by a two-tailed sign test where  $\alpha = \frac{.05}{.05}$ .

differed in type (one was related and the other unrelated). These results are summarized in Table 7. In general, the choices of related pairs occurred more often than expected (if choices are due solely to chance) at the lower actual frequencies and choices of related pairs occurred less often than expected (by chance) at the higher levels of actual frequency. This pattern of results is strongest for the subjects who studied the list under running frequency and degree-of-relation instructions, and the effect is somewhat attenuated for the subjects who studied by assessing background frequency. This is the same pattern as the one that occurred in the absolute estimates of intact pairs discussed above. Preference for related pairs at lower actual frequencies and preference for unrelated pairs at higher actual frequencies is exactly what would be predicted for this measure if subjects have frequency information that permits sharper frequency discrimination for unrelated pairs. If subjects were under the influence of a general bias, they should prefer related pairs at every level of actual frequency.

Next consider the frequency discrimination test in which the pairs compared were not of equal frequency. These data are summarized in Table 8. First note that the pattern of results for all study conditions is close to that obtained in Experiment 1 (see Table 2). In Experiment 7, however, error rates are lower than ever. The mean error score out of 6 is .33, and this rate is so low that the finer grained analysis conducted in Experiments 1 and 2 could not be useful in this experiment. Nevertheless, valuable conclusions can be drawn on the basis of the data in Table 8. A one-way analysis of variance [F(11,348) = 2.75, MSW = .277] was conducted on the number of errors for each

Table 8: Mean error scores out of 6 for the frequency discrimination test (unequal actual frequencies) of Experiment 7. (s.d.).

Comparison type	Higher frequency pair in comparison:		R	U	
	Lower frequency pair	r	u	r	u
	Degree of Ass'n	.37(.61)	.37(.55)	.13(.35)	.27(.45)
STUDY MODE	Background F'y:	.57(.68)	.27(.52)	.13(.35)	.20(.41)
	Running F'y:	.60(.77)	.50(.51)	.20(.41)	.30(.53)

subject for each condition shown and this analysis was used to assess a number of planned constrasts across the means of Table 8. First, no attributable to the different significant difference was study conditions. Second, as was found in Experiment 1, more errors occurred if the higher frequency pair in the comparison was related (critical difference = .109; obtained difference = .117). Third, also as found in Experiment 1 and in Experiment 2 after similarity instructions, fewer errors occurred in heterogeneous comparisons (Ru and Ur) than in homogeneous comparisons (Rr, Uu) (critical difference = .109; obtained difference = .117). Fourth, also as found in Experiments 1 and 2, more errors occurred in Rr comparisons than in Uu comparisons (critical difference = .154, obtained difference = .256). These latter two results were obtained in previous experiments but this is the first time the results are statistically significant. All these results are consistent with the view that the main difference between related and unrelated pairs lies in lower sensitivity or reduced discriminability (across levels of frequency) for related pairs relative to unrelated. These results are quite inconsistent with any view which asserts the existence of a bias favoring the over-estimation of the frequency of related pairs. The strongest prediction of such a bias view is that more errors should occur if the lower frequency pair in a frequency discrimination comparison is related. This was the final planned contrast conducted on these data and it did not approach statistical significance.

<u>Cued recall</u>. The final analyses concern the measures associated with the concurrent cued recall/frequency estimation task. First, consider the number correct in cued recall (out of 6) shown in Figure

Figure 15: Cued recall performance for subjects who had studied by assessing A, the semantic relationship; B, background pair frequency; or C, running frequency



MEAN PROPORTION RECALLED

This figure shows clearly that recall performance was much better 15. for related than for unrelated pairs. In fact, recall performance for related pairs is very close to ceiling in all instruction conditions. Because related pairs are close to perfect performance there is no room for improvement in recall over study frequency, while unrelated pairs do show an effect of actual study frequency. It is with unrelated pairs that most of the differences occur. The advantage for related pairs is greatest for subjects who performed running frequency estimation during study, and the advantage decreases in the other study conditions. This is basically a difference in performance for unrelated pairs because related pairs produce close to perfect recall under all study conditions. In addition, the improvement in recall for unrelated pairs with study frequency is greatest for those conditions in which recall performance is worst overall. All these observations are supported by an analysis of variance on the number of targets correctly recalled. The analysis included study condition as a between-subjects factor and item type and study frequency as within-subjects factors. This analysis yielded a main effect for study, indicating differences in overall recall probability; the means are .81, .74 and .52 for the degree-ofassociation, background frequency and running frequency conditions respectively [F(2,27) = 27.84, MSE = 1.78]. The analysis yielded a main effect for item type, indicating an advantage for related pairs [mean probabilities are .91 vs. .53, F(1,27) = 223, MSE = 1.48], and a main effect of study frequency [F(2,54) = 29.69, MSE = .622]. These main effects are qualified by interactions indicating the advantage for related items was greater in some study conditions than in others [F(2,27) = 8.24], indicating the advantage for related items is greater

at lower levels of actual frequency [F(2,54) = 3.86, MSE = .766], and indicating the advantage conferred by study frequency is greater in some study conditions than in others [F(4,54) = 4.89]. In general, a ceiling effect for related pairs prevented the observation of a simple overall advantage in recall performance for related pairs across levels of study frequency.

Accompanying frequency estimates. Next consider the frequency estimates produced in this cued-recall task. These data are shown in Figure 16. Generally, related pairs received higher overall estimates than did unrelated pairs, and the slopes of the estimation functions were steeper for related pairs, although this difference did not occur for subjects who studied by estimating on-line frequency. The overall magnitude of estimates also varied with study condition as did the steepness of the functions. That is, as expected, there was a positive relationship between the magnitude of estimates and their discriminability. Note that in the degree-of-association study condition, where the magnitude of estimates is the greatest (and where recall performance was the best), the related pairs actually show somewhat less accurate performance than unrelated pairs in the sense that mean estimates are closer to true values for unrelated items. These general conclusions were supported by analysis of variance of subjects' mean frequency estimates. As in previous analyses, instructions was a between-subjects factor and study frequency and pair type were within-subjects factors. The analyses yielded a main effect indicating that the magnitude of frequency estimates varied with study; means were 2.65, 2.24 and 1.97 for the degree-of-association, background frequency, and running frequency conditions, respectively [F(2,27) =

Figure 16: Mean frequency estimates accompanying cued recall for subjects who had studied by assessing A, the semantic relationship; B, background pair frequency; or C, running frequency



6.21, MSE = 1.12]. Significant main effects also indicated an overall advantage for related pairs [means = 2.70 vs. 1.88, F(1,27) = 65.34, MSE = .470], and the usual effect of actual frequency [F(2,54) = 191, MSE = .341]. Interactions qualified the main effects indicating sharper discrimination among frequency classes for related pairs [F(2,54) =23.26, MSE = .018] and differential discrimination for frequency depending on study instructions [F(4,54) = 5.30]. Finally, the discriminability advantage for related pairs did not occur following the running frequency study condition [F(4,54) = 56.69]. In general, the results of this task parallel the pattern for cued recall. When recall performance is very high, frequency estimates appear to reflect the operation of a bias for related pairs.

Finally, consider the frequency estimates conditionalized upon cued-recall performance, shown in Figure 17. Some of the points in this figure correspond to very few observations. Very few subjects fail to recall any targets from related pairs and, especially at lower levels of actual frequency, few subjects recall any unrelated targets. Nevertheless, the pattern of results for conditionalized estimates is quite different from the pattern for unconditionalized (Figure 16) estimates. As expected, frequency estimates for correctly recalled pairs are higher, more accurate and show greater discriminability than estimates for pairs whose targets are not recalled. The advantage for related pairs in magnitude and accuracy evidenced in unconditionalized results is eliminated and perhaps even reversed when frequency estimates are conditionalized on correct recall. Analyses of variance on the estimates for correctly recalled pairs supported some of these conclusions. Analyses were conducted only for the study conditions in

## Figure 17:

Mean frequency estimates conditionalized on cued recall for subjects who had studied by assessing A, the semantic relationship; B, background pair frequency; or C, running frequency



ACTUAL FREQUENCY

which the subjects studied by assessing degree of association or background frequency because too few subjects in the running frequency condition recalled any unrelated items. In both cases there was a main effect for frequency [F(2,18) = 185.24, MSE = .220 and F(2,18) = 69.64,MSE = .471, respectively], and no effect for pair type. In the case of the background-frequency condition, the analysis yielded a reliable interaction, indicating higher discriminability (a steeper slope for estimates across actual frequency) for unrelated pairs [F(2,18) = 5.97,MSE = .247). As in the case of unconditionalized estimates, the degreeof-relation study condition produced generally inflated estimates and the estimates for unrelated pairs are actually closer to true frequency values. Although an analysis of variance was impossible because too few subjects could be included, inspection of the data in Figure 17 for the running-frequency groups shows clearly that the estimates for unrelated pairs show more discriminability and accuracy than the estimates for related pairs.

<u>Comparisons among measures</u>. Now consider the differences between the measures used in this experiment. First of all, note that despite very accurate performance, the typical slope difference between the estimates for related and unrelated pairs is obtained even in continuous estimation. In this case, subjects are studying all pairs in the same way - by estimating on-line frequency. This result seems to cast doubt upon the theory that the reliable differences in slope or intercept obtained throughout this research result from differences in the way in which subjects deal with the experimental pairs at the time of final test, as suggested by the postulation of a response bias. The assertion that subjects base their estimates on the associative connection between pair members (Tversky & Kahneman, 1973) suggests that a bias operates that the time of test. Recall too, that in Experiment 1, subjects engaged in frequency estimation during study. Note that the significant interaction (see Figure 12) stems almost entirely from the difference between the two classes of items at the fourth presentation. That is, no imporant differences in accuracy or slope occur until after several presentations. This result is quite consistent with the utility-of-trace-information argument presented before. It is not unreasonable to postulate that, although subjects engage in on-line frequency estimation, they notice semantically associated pairs when they occur. Noticing and encoding information relevant to semantic characteristics should result in traces of lower utility in frequency estimation even a short time later.

This postulation seems to run into some trouble, however. According to the theory, those subjects who study the list by estimating on-line frequency should be engaged in a process that accentuates or focuses upon the episodic nature of the study task. This should be especially true in comparison to the subjects who engage in a task that focuses upon the semantic characteristics of the study material, as do the subjects who study by assessing the degree of relationship between the study pair members. Thus on-line frequency study should result in more accurate performance on a frequency estimation test than should the degree-of-association task. Although the on-line estimates are remarkably accurate, the final test performance for this on-line frequency group on almost all measures shows lower accuracy than the performance of the other two study groups. That is, this episodic study task does yield more accurate performance but the accuracy is not

maintained over time. It appears that the on-line frequency task results in poorer retention of the study materials, as shown by the cued-recall task. This poorer retention, in itself, should lead to estimates of lower magnitude and slope and frequency discrimination of lower accuracy.

In general, the frequency discrimination measures support the same conclusions as the frequency estimates for intact pairs. That is, these data show that the main difference between related and unrelated pairs lies in the accuracy of discrimination with which subjects can make frequency judgments. Both measures indicate that this result is attenuated or even reversed for subjects who study the list by assessing the frequency with which pair members occur together in everyday experience. The only difference between these two measures is that the frequency estimation measure shows significant differences between the instructional groups in the discriminability or slope of the estimation functions. Both measures show that the running frequency study group performed with less accuracy than did the other two groups, but this difference is not significant for the frequency discrimination measure, probably because performance for all groups was so near perfect.

As predicted, subjects who study by assessing background frequency do not show slope, discrimination, nor accuracy differences between related and unrelated pairs. Theoretically, attending to background frequency equalizes the episodic/semantic content of the memory traces for related and unrelated pairs and thereby produces frequency judgments of more or less equivalent accuracy.

Finally, the measures of frequency estimation accompanying cued recall and frequency estimates for single items show very similar

This pattern is quite different from that obtained in patterns. frequency estimates of intact pairs. This difference occurs for both measures for the same reason. Because frequency information accrues to the pair as a memory unit, the retrieval of the whole pair is necessary for successful frequency estimation of even single items. Single items or recall cues that come from related study pairs are more likely to permit retrieval of the entire pair than are items from unrelated pairs. Therefore single words or cues from related study pairs yield higher and steeper frequency estimation functions. The same pattern is obtained for all study instructions except that, for both measures, estimates appear to be lower in magnitude and slope for the on-line frequency group overall. This result is consistent with the foregoing argument as long as it can be assumed that this frequency estimation study task yields poorer overall retention than do the other, more semantic, study tasks. The data from the cued-recall measure appear to support this assumption. In sum, these two measures of frequency estimation yield results that look very much like a bias favoring the relative overestimation of related pairs. This result is obtained only because intact pairs are not presented for estimation.

The conclusions drawn about the difference between frequency judgments for related and unrelated events depend very much upon the measures employed and the ranges of actual frequency sampled. It is of interest to speculate on the results obtained and conclusions that would be drawn if still other measures were employed in this kind of experiment. For example, it is well established that frequency estimates provide levels of old/new discrimination equivalent to tests of recognition (Flexser & Bower, 1975; Malmi, 1977; Harris et al., 1980). It is usually supposed that this is because recognition and frequency estimation are accomplished with reference to the same memory information (cf. Harris et al., 1980). If this is the case, recognition tests of intact pairs, following a semantic study task, would show better old/new discrimination for unrelated than for related pairs. The opposite results should be obtained if subjects must perform a recognition test on single members of the study pairs. That is, under single testing, members of related study pairs should show better old/new discrimination than members of unrelated study pairs because related members have a higher probability of reinstating the study context. Humphreys (1978) has reported this same result.

## Conclusions

Overall, the goals of this experiment have been met. The potential for intralist confusion for unrelated pairs has been removed and the conclusions of the first experiments have not been qualified. The frequency estimates for individual pair members and for study pairs on-line have behaved precisely in the ways predicted. Permitting overestimation at all levels of actual frequency provides evidence that accuracy accompanies discriminability. In general, increases in slope are accompanied by increases in average accuracy. Finally, the comparison of study groups supports the contention that even when instructed to attend to frequency, subjects' typical encoding strategy involves attention to the very salient semantic association between members of related pairs. The difference in the episodic/semantic nature of the encoding of related and unrelated pairs can be removed, however, if subjects are required to assess background frequency of the pair.

As a final point, consider the effect of instructions upon frequency estimates. As mentioned before, the consistent lack of any effect of study instructions to attend to frequency (as opposed to memory instructions) has been interpreted as evidence that the accrual of frequency information is an automatic process (Hasher & Chromiak, 1977; Hasher & Zacks, 1979). There are, however, several manipulations of study instructions that do have effects on the accuracy of frequency For example, subjects who study a list under a semantic estimates. instruction (Rowe, 1974) or imaginal set (Rowe & Rose, 1977) or by pronouncing the items (Hopkins et al., 1972) produce frequency estimates that are higher in magnitude and more accurate than estimates by subjects who do not engage in these forms of study. Experiment 7 provides evidence that subjects who study under running frequency. instructions produce frequency estimates that are less discriminating among levels of actual frequency than subjects who study under other instructions. In addition, of course, related and unrelated pairs yield differences in frequency discriminability. The episodic/semantic encoding theory developed in this thesis asserts that the accrual of frequency information is not equivalent and automatic for all events in all contexts, and the data provided by this experiment are inconsistent with any theory that states that the accrual of frequency information is automatic in the usual sense of that term. The automaticity view, on the other hand, (Hasher & Zacks, 1979) is in the difficult position of explaining why an automatic process is more automatic for some kinds of events or some study instructions than for others. It seems very likely that those variables that do not affect frequency estimation also will not affect other measures of memory retention (especially

recognition). The relegation of frequency estimation to the status of "automaticity" denies ample evidence that frequency judgment is a measure of memory retention interrelated with other measures - most especially recognition.

## Summary, Conclusions and General Discussion

The first goal of this thesis has been to establish a basic empirical relationship between frequency estimates for related and unrelated events. In standard situations involving absolute estimation, relative frequency judgment, or judgments of correlation, the empirical relation is best described by saying that subjects have more sensitive, more discriminating and, in general, more accurate information about the frequency of unrelated pairings than about related. The introduction to this thesis was intended to show that this empirical relationship is quite consistent with what else is known about frequency judgments.

The theoretical position proposed in this thesis makes a number of simple assumptions in order to account for this empirical relationship. First of all it is necessary to distinguish between two hypothetical sorts of memory information that a subject can have about an event. The first is the ability to produce the event or to generate a token of the event type. The second is the ability to make a differential response to old events versus novel events. These two memory abilities are most closely associated with measures of recall and recognition respectively, although it is not necessary to assume that these two measures exactly assess the two hypothetical memory abilities. Similarly, it is not necessary to assume that the two abilities or measures are completely independent, merely that sometimes they deviate from perfect dependency. Other assumptions of this account are that subjects try to attend to the associative connection between pair members, and in the case of related pairs, this attention results in a relatively general, semantic memory trace. In the case of unrelated

pairs, attention to the associative connection results in a relatively episodic trace. It is also assumed that a memory trace containing relatively episodic information is much more useful in a frequency judgment task (involving intact study pairs, at least) than is a trace containing relatively semantic information. An episodic trace contains information specific to the time and place of an event's occurrence while a semantic trace does not. Old/new discrimination can only be accomplished with reference to the first sort of memory information, and old/new discrimination is a necessary first step to frequency judgment. The more useful is a 'set of information, the more precise and discriminating are judgments based on that information.

Not all investigators agree with this characterization of the empirical or the theoretical relationship, however. Chapman (1967) and Tversky and Kahneman (1973) have described the difference between subjects' frequency judgments for related and unrelated pairings as a bias toward the relative overestimation of related pairs. In both cases, the authors account for this bias in terms of the associative connection between members of related pairs. That is, subjects are assumed to assess the strength of the associative connection during test and then produce frequency estimates that reflect the strength of the associative connection rather than relevant memory information about the study phase. It follows, then, that subjects will make errors when and if the associative connection is not perfectly related to situational The approach of Chapman (1967) and Tversky and Kahneman frequency. (1973) differs from the view proposed in this paper in some fundamental ways. First, the theory of these authors is basically not an episodic memory theory. Subjects base their decisions on a judgment heuristic

and not directly upon what is retained about the study phase. There are undeniably situations in which subjects must have little or no memory information about test events, especially events that were not attended In such circumstances, it seems reasonable to suppose subjects will to. base their decisions on whatever information they have about the test items. However, it seems unlikely that subjects who are informed about the nature of the test (as were Chapman's, 1967) would not possess or would ignore relevant memory information that they could use directly at test. A second related difference between the approaches is that the associative-connection account seeks to explain the difference between related and unrelated pairs in terms of differences in the subjects' behavior during the test. Conversely, the episodic/semantic encoding approach explains the related/unrelated difference in terms an interaction between subjects' behavior while the study events occur and the nature of the test. This contrast between approaches has implications for techniques one might recommend to improve people's For example, the episodic/semantic encoding frequency judgments. approach might recommend that subjects deal with events in terms of episodic characteristics regardless of how expected the events seem to If the problem is one of a bias, however, it could be recommended be. that subjects intentionally underestimate the frequency of expected co-Clearly, neither proposed course of action would be occurences. expected to work by the opposing view.

In any case, the data presented in this thesis permit the conclusion that the empirical relationship proposed by Chapman (1967) and Tversky and Kahneman (1973) is simply wrong. In standard situations with standard instructions, subjects do not behave as if they have a

bias toward the relative overestimation of the frequency of related pairs. This erroneous conclusion may have been due to two factors. First, Tversky and Kahneman (1973) used word pairs that were related and unrelated in a sense different from that employed in the research reported here. In that experiment a pair of personality characteristics was called "related" if pilot subjects felt both characteristics were likely to be true of one person. Thus, a pair like happy-sad would probably be treated as unrelated by the Tversky and Kahneman (1973) procedure while happy-sad would be a related pair if used in the experiments reported in this thesis. Second, and more important, much (not Tversky & Kahneman, 1973) of the research upon which the bias conclusion was based did not employ a full range of actual frequencies. It is important to note that at an actual frequency of zero, imprecision in frequency information must result in estimates that are higher than actual values because subjects cannot give estiamtes lower than zero. Thus at relatively low actual frequencies, subjects will probably give higher estimates to that class of items about which they are more prone to make errors. If an investigator only considers relatively low levels of actual frequency, he/she may obtain results that look as if subjects have an overall response bias when, in fact, classes of items differ in the discrimination with which subjects make decisions about frequency.

The research reported in this thesis shows that there are a number of other nonstandard situations in which subjects will behave in ways that could be described in terms of a bias. Most important, if the frequency judgment task depends primarily on the memory ability associated with production or generation of the event, the associative connection between pair members will confer an advantage on the related

pairs. That is, frequency information accrues to the study event as a whole. Subjects cannot effectively estimate frequency unless they can reinstate the study event. The associative connection between members of semantically related pairs increases the probability that the entire pair will be reinstated given one member, perhaps because subjects need entertain only a small number of candidate targets in order to generate the correct target. Therefore, frequency estimates that accompany cued recall and estimates of individual pair members reflect recall performance which in turn reflects the strength of the associative connection between pair members. For these measures, then, frequency estimates are often higher for related pairs than for unrelated. These measures also produce more discriminability for related items, because frequency information accrues to the entire study pair.

Some instructions remove discriminability differences between related and unrelated pairs and may even produce results that could be described as a bias. If subjects change their study strategy, they change the way in which they operate on the study material. Attention to contrasts between pair members, and attention to background frequency of the pair equalize the episodic/semantic content of the encoding operations and the resultant memory traces for related and unrelated pairs. When memory traces for related and unrelated pairs. When memory traces for related and unrelated pairs are equivalent in relative episodic and semantic content, subjects make frequency judgments that are equally discriminating for the two classes of items. In such cases, large differences in retrieval probability (even when intact pairs are tested) seem to result in a slight bias toward the relative overestimation of related pairs.

In general, in standard circumstances, related and unrelated pairs differ in discriminability and accuracy, not in overall level. If certain nonstandard measures, special instructions, or restricted actual frequencies are used, however, subjects will provide results that could be described as a bias.

Relation to Traditional Theories of Frequency Estimation

As mentioned in the introduction, the traditional theories of frequency judgment agree that frequency estimates depend upon old/new discrimination or recognition performance. Hintzman and Block's (1971) multiple-trace theory states that each occurrence of an event establishes a separate memory trace. Recognition requires contact with an appropriate trace, then frequency judgments require the additional step of estimating the number of such traces. Similarly, each occurrence of an event can establish a list marker (Anderson & Bower, 1972) at the permanent address of the item. Recognition requires the discovery of at least one list marker, and frequency judgments require an estimate of the number of markers. Likewise, Underwood (1969) assumes that the memory trace for an item is a bundle of attributes, one of which is specifically sensitive to event frequency. Underwood asserts that recognition decisions are accomplished exclusively by reference to the frequency attribute (Underwood et al., 1971). Traditional theories agree on the importance of old/new discrimination for frequency judgments, but do not provide a basis for deciding whether the accuracy of this old/new discrimination will differ for related and unrelated pairings. The episodic/semantic encoding view postulated in this thesis provides such a basis.

Not all these classical theories of frequency estimation can adopt the episodic/semantic encoding postulate equally easily, however. Hintzman and Block (1971) assert that the memory traces are identifiable by time tags. However, the statement that the memory traces for some items possess more accurate temporal (episodic) information than others necessitates the qualification of that multiple-trace postulate to permit the differential loss of discriminability among traces over time. Attribute theory, on the other hand, makes no assertions about the temporal independence of frequency information - in fact, the simplest version would assert that there is no temporal information represented with frequency - they are separate attributes or components of the memory trace. This simplest version of attribute theory must assume that the frequency counter for unrecognized items is zero, and therefore has difficulty accounting for the fact that overall magnitude differencs often do not accompany slope or discriminability differences. Theoretically, old/new discrimination suffers because frequency attributes are lost, or are not updated during presentation. Frequency estimates for items that receive less discriminating estimates should also be lower. It has been suggested (Begg, 1974) that subjects adjust their estimates with reference to the overall task mean. However. attribute theory asserts that frequency information is specifically Independent knowledge about the task mean would seem to encoded. require an attribute for separate classes of events. In any case, theories that assert that frequency estimates come from estimates based on more general memory information more naturally permit subjects to retain some idea of the relative frequency for classes of items and adjust their final estimates accordingly.

This issue of explicit versus inferred frequency information cuts across the traditional theories of frequency estimation. Hintzman (1976) refers to this distinction as multiple traces versus propositional encoding. Hintzman states that one difference between these two general approaches is that "multiple-trace hypothesis assumes that the subject determines frequency at the time of [test], while the propositional encoding hypothesis assumes that the subject encodes frequency information while studying the list (p. 59)." Note that this does not necessarily imply that the multiple-trace hypothesis asserts that the information used to make frequency judgments does not accrue during study, but that the information that accrues during study is not specifically frequency information. In fact, the multiple-trace hypothesis asserts that the accrual of this general information, later used for frequency judgments, must accrue during study. Because memory traces are the basic fabric of the memory system, frequency information is always present. Propositional approaches permit the subject to fail to encode specific frequency information (Hintzman & Stern, 1978). Unlike multiple-trace theory, at least one theory that asserts that frequency judgments are inferred from other memory information can be rejected. Based on the data presented in this thesis, it seems fair to conclude that subjects usually do not infer frequency judgments directly from the semantic, associative connection between pair members. In general, however, this distinction between inference and propositional encoding may provide a means to usefully combine the traditional theories of frequency presentation.

The results of Experiments 2 and 4 suggest that discrimination on the basis of semantic characteristics is accomplished somewhat independently from discrimination on the basis of frequency. The similarity/contrast manipulation has been shown to produce differences in recognition and recall only when the study instructions are relevant to the discrimination necessary in the memory task (Begg, 1978). The results of Experiment 2 show that the semantic discrimination made during study can be largely irrelevant to the accuracy of frequency discriminations required during the test. The tentative conclusion to be drawn is that semantic and frequency (episodic) aspects of the memory trace can operate quite independently. This independence of memorial attributes is a basic assumption of one propositional encoding theory. It follows naturally from attribute theory (Underwood, 1969) that decision on the basis of semantic (or associative) attributes need have no necessary implications for discrimination on the basis of frequency (episodic) attributes. Thus attribute theory conveniently captures the notion that the memory trace for an item or event is not a unitary, allor-nothing entity. By understanding the interaction of test and study in terms of the relative nature of trace composition, it is possible to assert that the usefulness of a trace or traces depends upon the relationship between study and test (Watkins & Tulving, 1975; Morris, Bransford & Franks, 1977). An encoding of a single memory trace involves the selection, by the study task, of some subset of the possible features that could be encoded. If the test task makes use of the encoded features, performance will be optimal. If, however, the test requires the use of a feature that was not initially encoded, performance will suffer. Such an account can still assume a separate

trace for every study episode so that the usefulness of the trace contact aspect of multiple-trace theory is retained. The notion that each individual trace is a collection of features and that these features can correspond to any encoding procedure is taken from attribute theory. In fact, this modification or extension of multipletrace theory can be viewed as a fruitful marriage or reconciliation between the attribute and multiple-trace approaches to frequency theory. This marriage must, of course, imply that frequency information is both explicitly and implicitly coded. In a very real sense this must be There can be little doubt that frequency information can be true. explicitly coded and retained - people do count and remember totals. Likewise, it is highly likely that people could estimate the frequency for events which had not been coded per se. Unless one wishes to postulate a separate freuency attribute and memory trace for every possible test item (e.g., the frequency of the letter 'n', say) it seems undeniable that people can estimate frequency using inferences based on more general memory information. Instead of competition between theories, a more useful enterprise would be the investigation of the domain of operation for the two presumed strategies.

Clearly, this combination of traditional theories fits well with the episodic/semantic encoding view postulated on this thesis. It is quite reasonable to assume a multiplexing of the information encoded in a memory trace. That information can involve semantic, associative attributes or episodic attributes. Clearly, episodic information is much more useful in a later test of situational frequency. This episodic/semantic encoding approach is not the only way to explain the

results reported in this thesis. I will now briefly consider other theoretical statements consistent with the data.

## Other Theoretical Approaches

The episodic/semantic account postulated in this thesis explains frequency discrimination differences in terms of differences in the quality and usefulness of memory traces. Logically, it is possible to propose an explanation that relies upon presumed differences in the quantity of memory information. Recent theoretical work within the levels-of-processing framework has explained memory performance in terms of differences in the elaboration of memory traces (Craik & Tulving, 1975). By this view, elaboration or "spread" of encoding is a dimension somewhat orthogonal to the quality or level of processing. It could be postulated that, within the semantic level, related pairs receive relatively superficial, unelaborated processing during study because the relationship between pair members is routine and expected. Unrelated pairs, on the other hand, receive deeper processing because of their uniqueness and, consequently, unrelated pairs are better remembered and receive more accurate and more discriminating frequency estimates. Related pairs are better remembered in cued-recall, however. This reversal might be explained by saying that related pairs enjoy an advantage in the restriction of candidate targets subjects must entertain. In support of this contention, one can point to the higher frequency estimates for unrelated pairs after estimates are conditionalized upon recall performance. In general, a levels-ofprocessing approach, in terms of differences in quantity of memory information rather than quality, is not at odds with the episodic/semantic encoding view postulated here. The episodic/semantic

approach, however, more naturally captures the important notion of the relativity of memory information. The value of memory traces does not lie in their overall goodness, strength or depth, assessed independently. Instead, the utility of memory information depends upon the relationship between the information retained and the requirements of the particular task of interest. These ideas are captured in more recent theoretical work within the levels-of-processing framework. It has been suggested, for example, that within the semantic level, the amount of processing, relevant to the memory task, is crucial for later performance (Johnson-Laird; Gibbs, & de Mowbray, 1978). Similarly, (Jacoby, Craik & Begg, 1979) see no point in explaining variations in performance within the semantic level in terms of sublevels that vary in the degree of detailed analysis. Instead, these authors propose the idea of distinctiveness. Unlike elaboration, distinctiveness is relative rather than absolute. Thus the importance of differences in processing lies not in the depth of encoding but in the "formation of more precise descriptions, hence a more distinctive memory record." By this view, the distinctiveness of the memory trace is especially important for tasks involving discrimination of traces from each other and background noise. Pre-experimental strength and reconstructive factors of the context are especially important for recall.

Another approach to explaining differences in frequency discriminability in terms of quantitative differences in the representation of frequency comes from work on verbal discrimination learning briefly mentioned in the introduction (Goedel & Thomas, 1977; Eckert & Kanak, 1974). For the purposes of this thesis, the important aspects of the work on verbal discrimination learning are the research

and theory provided on the issue of how variables affect subjective frequency. The basic conclusion that can be drawn from that work is that increments in subjective frequency may not depend on the way frequency is manipulated. It appears that any manipulation which causes subjects to treat one class of items differently than another produces results that are interpreted as the differential accrual of subjective frequency. Thus familiarization with some items before study, varying situational or background frequency, or manipulating feedback can all apparently influence the accrual of subjective frequency. Frequency theory is often assumed to include a postulate analogous to the Weber law. Thus, the addition of a unit of frequency to an item already high in subjective frequency produces a smaller perceptual difference than the addition of a frequency unit to an item of low subjective frequency. This Weber law is consistent with the typical negatively accelerated shape of many frequency estimation functions. It is also consistent with the observation that items of relatively high subjective frequency (e.g., abstract words) show worse frequency discrimination performance than items of relatively low subjective frequency (cf. Galbraith & Underwood, 1973; Begg, 1974).

Furthermore, such a principle would predict that when subjects study items of varying background linguistic frequencies, if there is any dependency in the session-to-session representation of frequency information (however conceived), items of high background frequency will show worse frequency discrimination than low frequency items (Reichardt et al., 1973). Nevertheless, it can be expected that the items of high background frequency would be better recalled. This relationship has been demonstrated for recognition (Shepard, 1967), and recall (Hall, 1954) and if related pairs can be viewed as more frequent, familiar events than unrelated pairs, this Weber law postulate extends easily to the experiments reported here. This extension does require some statement about why discriminability differences are observed without differences in the overall magnitude of estimates. Apparently it is still necessary to assume that subjects retain some notion of the relative frequency of the two classes of study items and use this information to adjust their frequency judgments accordingly.

The Weber law is valuable because it suggests that frequency discriminations are based upon relative rather than absolute frequency differences (Goedel & Thomas, 1977). One apparent problem with this approach seems to be the implication that a manipulation has its primary effect in changing subjective frequency. That is, it seems to be assumed that some memory tasks are accomplished with reference to explicitly stored frequency information and that this stored frequency information is retained in some unitary undifferentiated way. The assertion that a variety of manipulations operate to affect the strength of a single memorial representation has often been discounted (e.g., Hintzman, 1976). The fact that subjects can remember exposure duration and recency information independently of frequency information suggests that, contrary to a strength hypothesis, subjects have access to information about individual presentations and that the effects of one presentation can be discriminated from the effects of another. Although it may be possible to show that the variables that are assumed to increment subjective frequency by frequency theory do lose their identity after having their effects, I would be more comfortable with a

memory theory that did not place the explicit representation of frequency information in such a fundamental role.

Many of the theoretical ideas contained within the qualitative and quantitative accounts above can be subsumed by a more general approach to memory discrimination. Discussed briefly in the introduction, this more general approach can subsume the episodic/semantic account proposed in this thesis by stating that subjects deal with related and unrelated pairs in different ways. The related pairs receive processing that is routine and predictable while the unrelated pairs receive processing unique to the experiment (mainly because they are almost entirely novel events). At test, the memory trace for related pairs contains information that is of little value in discriminating these related items from background frequency because study processing has emphasized routine aspects the related pairs share with background presentations. Almost by definition the remnants of the processing of unrelated pairs are of more value in permitting discrimination from background. By this account, the difference in discrimination from background results in a difference in situational discrimination and the discriminability of frequency judgments. Up to this point this general approach is isomorphic with the episodic/semantic account. The general approach can go further, however, by asserting that any difference in the kind or amount of processing should produce the same results. This general approach is not committed to explaining all frequency discriminability differences in terms of the episodic/semantic content of memory traces. For example, the differential frequency discrimination for underlined and not underlined words can be understood as the result of a change in the
typical or routine processing for underlined items (Radtke, Jacoby & Goedel, 1971). This approach may be able to explain a wider range of phenomena by postulating changes in the quality <u>or</u> quantity of processing from typical or routine levels. As such, the episodic/semantic account proposed here can be thought of as a subset of this wider view. The research report here establishes the value of the episodic/semantic account for an understanding of judgments of frequency and correlation. A determination of the value of this broader theoretical approach must, however, await considerable further research.

## Predictions and Extensions

Suggested further research would be addressed to the question of whether the empirical results and proposed explanations discussed in this thesis can be generalized to other situations, materials and meaures. For example, Experiment 5 established that the basic empirical relationship extended to judgments of conjoint frequency. The theory proposed in this thesis would predict that the same basic relationship should apply to judgments of correlation. As another example, Tversky and Kahneman (1973) did not obtain the same discriminability advantage for unrelated pairs that is consistent throughout this thesis. This difference in obtained results is probably due to the operational definition of "related." It may be that the empirical relationship reported here is very sensitive to the choice of experimental materials. As another example, the viewpoint proposed in this thesis asserts that for frequency judgments involving intact pairs, generation or recall of the study event(s) is relatively unimportant and old/new differentiation is crucial. It is quite conceivable, however, that real world situations involve relatively complex, variable "study" events. If the judgment of frequency or correlation depends upon the implicit or explicit presence of the encoding episode, the presentation of two words, labels, ideas, etc., may not be sufficient to reinstate the encoding events. If this were true, subjects might indeed display a bias towards the relative overestimation of related events. Thus. confident extensions of the results and theory of this thesis to other contexts must await assessment of questions like these.

127

Ì.

Nevertheless, predictions in the area of human judgment and decision making can be made, based on the arguments presented here, and the results of these experiments can be generalized to situations in which people must make judgments of correlation. In such situations, it has been reported that people are subject to illusory correlation. Items or ideas that have previously occurred together, or that seem as if they should go together, have been reported to show overestimation of contingency. If the results from this thesis are generalized to such situations one should not necessarily expect an overall overestimation of the correlation of related things. Rather, one should expect differences in the accuracy with which people can estimate the correlation between apparently related and apparently unrelated events. This difference in accuracy could well be expected to produce an overestimation for related events at relatively low levels of actual correlation. However, if the arguments presented here are correct, one should also find that when the actual correlation is high, the contingency between apparently related events should be underestimated relative to unrelated events. Physicians might, for example, actually underestimate the degree to which a particular symptom is an indicator of a particular disease, when that symptom is, in fact, a very good indicator. This underestimation should occur more when the relationship between the symptom and the disease is obvious or expected (though perhaps not perfect) as opposed to symptoms and diseases whose relationships would be considered unexpected or counterintuitive. This finding of a difference in accuracy for judgments of correlation might help to explain any apparently unwarranted degree of caution or

1

conservatism in the diagnosis and treatment of apparently straightforward cases.

## REFERENCES

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

- Battig, W.F., & Montague, W.E. Category norms for verbal items in 56 categories. Journal of Experimental Psychology Monographs, 1969, 80, (3, Pt.2).
- Begg, I. Estimation of word frequency in continuous and discrete tasks. Journal of Experimental Psychology, 1974, 102, 1046-1052.
- Begg, I. Similarity and contrast in memory for relations. <u>Memory and</u> Cognition, 1978, 6, 509-517.
- Begg, I., & Rowe, E.J. Continuous judgments of word frequency and familiarity. Journal of Experimental Psychology, 1972, <u>95</u>, 48-54.
- Bourne, L.E., Ekstrand, B.R., Lovallo, W.R., Kellog, R.T., Hiew, C.C., & Yaroush, R.A. Frequency analysis of attribute identification. Journal of Experimental Psychology: General, 1976, <u>105</u>, 294-312.
- Chapman, L.J. Illusory correlation in the observational report. Journal of Verbal Learning and Verbal Behavior, 1967, <u>6</u>, 151-155.
- Chapman, L.J., & Chapman, J.P. Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. <u>Journal of Abnormal</u> Psychology, 1969, 74, 271-280.
- Craik, F.I.M., & Tulving, E. Depth of processing and the retention of words in episodic memory. Journal of Experimental Psychology: General, 1975, 104, 268-294.

- Eckert, E., & Kanak, N.J. Verbal discrimination learning: A review of the acquisition, transfer, and retention literature through 1972. Psychological Bulletin, 1974, 81, 582-607.
- Ekstrand, B.R., Wallace, W.P., & Underwood, B.J. A frequency theory of verbal discrimination learning. <u>Psychological Review</u>, 1966, <u>73</u>, 566-678.
- Estes, W.K. The cognitive side of probability learning. <u>Psychological</u> Review, 1976, 83, 37-64.
- Flexser, A.J., & Bower, G.H. Further evidence regarding instructional effects on frequency judgments. <u>Bulletin of the Psychonomic</u> <u>Society</u>, 1975, <u>6</u>, 321-324.
- Galbraith, R.C., & Underwood, B.J. Perceived frequency of concrete and abstract words. <u>Memory and Cognition</u>, 1973, <u>1</u>, 56-60.
- Ghatala, E.S., Levin, J.R., & Wilder, L. The apparent frequency of words and pictures as a function of pronunciation and imagery. Journal of Verbal Learning and Verbal Behavior, 1973, <u>12</u>, 85-90.
- Ghatala, E.S., Levin, J.R., & Wilder, L. Pronunciation and the frequency-meaningfulness effect in childrens' frequency discrimination. <u>Journal of Experimental Psychology: Human</u> <u>Learning and Memory</u>, 1975, <u>1</u>, 655-659.
- Goedel, G.D., & Englert, J.A. Imagery and phenomenal background frequency in verbal discrimination learning. <u>Memory and</u> <u>Cognition</u>, 1978, <u>6</u>, 209-216.
- Goedel, G.D., & Thomas, C.A. The imagery effect and frequency discrimination. <u>Memory and Cognition</u>, 1977, <u>5</u>, 53-58.
- Golding, S.L., & Rorer, L.G. Illusory correlation and subjective judgment. Journal of Abnormal Psychology, 1972, 80, 249-260.

- Hall, J.F. Learning as a function of word-frequency. <u>American Journal</u> of Psychology, 1954, 67, 138-140.
- Harris, G., Begg, I., & Mitterer, J. On the relation between frequency estimates and recognition memory. <u>Memory and Cognition</u>, 1980, 8, 99-104.
- Hasher, L., & Chromiak, W. Processing of frequency information: An automatic mechanism? <u>Journal of Verbal Learning and Verbal</u> <u>Behavior</u>, 1977, <u>16</u>, 173-184.
- Hasher, L., Goldstein, D., & Toppino, T. Frequency and the conference of referential validity. Journal of Verbal Learning and Verbal Behavior, 1977, 16, 107-112.
- Hasher, L., & Zacks, R.T. Automatic and effortful processes in memory. Journal of Experimental Psychology: General, 1979, <u>108</u>, 356-388.
- Hintzman, D.L. The effect of repetition and exposure duration on memory. Journal of Experimental Psychology, 1970, 83, 435-444.
- Hintzman, D.L. Repetition and memory. In G.H. Bower (ed.) <u>The</u> <u>Psychology of Learning and Motivation</u>, vol. 110, New York: Academic Press, 1976.
- Hintzman, D.L., & Block, R.A. Repetition and memory: Evidence for a multiple-trace hypothesis. Journal of Experimental Psychology, 1971, 88, 297-306.
- Hintzman, D.L., & Stern, L.D. Contextual variability and memory for frequency. <u>Journal of Experimental Psychology: Human Learning</u> and Memory, 1978, 4, 539-549.

Hopkins, R.H., Boylan, R.J., & Lincoln, G.L. Pronunciation and apparent frequency. <u>Journal of Verbal Learning and Verbal Behavior</u>, 1972, 11, 105-113.

- Howell, W.C. Storage of events and event frequencies. Journal of Experimental Psychology, 1973, 98, 260-263.
- Humphreys, M.S. Item and relational information: A case for context independent retrieval. Journal of Verbal Learning and Verbal <u>Behavior</u>, 1978, <u>17</u>, 175-189.
- Jacoby, L.J., Craik, F.I.M., & Begg, I. Effects of decision difficulty on recognition and recall. <u>Journal of Verbal Learning and</u> <u>Verbal Behavior</u>, 1979, <u>18</u>, 585-600.
- Johnson, M.K., Taylor, T.H., & Raye, C.L. Fact and fantasy: Effects of internally generated events on the apparent frequency of externally generated events. <u>Memory and Cognition</u>, 1977, <u>5</u>, 116-122.
- Johnson-Laird, P.N., Gibbs, G., & de Mowbray, J. Meaning, amount of processing and memory for words. <u>Memory and Cognition</u>, 1978, <u>6</u>, 372-375.
- Leicht, K.L. Recall and judged frequency of implicitly occurring words. Journal of Verbal Learning and Verbal Behavior, 1968, 7, 918-923.
- Malmi, R.A. Context effects in recognition memory: The frequency attribute. Memory and Cognition, 1977, 5, 123-130.
- Mathews, R.C. Semantic judgments as encoding operations: The effects of attention to particular semantic categories on the usefulness of interitem relations in recall. <u>Journal of Experimental</u> <u>Psychology: Human Learning and Memory</u>, 19777, <u>3</u>, 160-173.

- Morris, C.D., Bransford, J.D., & Franks, J.J. Levels of processing versus transfer appropriate processing. <u>Journal of Verbal</u> Learning and Verbal Behavior, 1977, 16, 519-534.
- Neumann, P.G. Visual prototype formation with discontinuous representation of the dimensions of variability. <u>Memory and</u> Cognition, 1977, 5, 187-197.
- Palermo, D.S., & Jenkins, J.J. <u>Word Association Norms</u>. Minneapolis: University of Minnesota Press, 1964.
- Proctor, R.W., & Ambler, B.A. Effects of rehearsal strategy on memory for spacing and frequency. <u>Journal of Experimental Psychology:</u> Human Learning and Memory, 1975, 1, 640-647.
- Radtke, R.C., Jacoby, L.L., & Goedel, G.D. Frequency discrimination as a function of frequency of repetition and trials. <u>Journal of</u> <u>Experimental Psychology</u>, 1971, 89, 78-84.
- Reichardt, C.S. Shaugnessy, J.J., & Zimmerman, J. On the independence of judged frequencies for items presented in successive lists. <u>Memory and Cognition</u>, 1973, 1, 149-156.
- Rose, R.J. Encoding variability, levels of processing, and the effects of spacing of repetitions upon judgments of frequency. <u>Memory</u> <u>and Cognition</u>, 1980, <u>8</u>, 84-93.
- Rose, R.J., & Rowe, E.J. Effects of orienting task and spacing of repetitions on frequency judgments. <u>Journal of Experimental</u> <u>Psychology: Human Learning and Memory</u>, 1976, 2, 142-152.
- Rowe, E.J. Frequency judgments and the recognition of homonyms. Journal of Verbal Learning and Verbal Behavior, 1973, <u>12</u>, 440-447.

- Rowe, E.J. Depth of processing in a frequency judgement task. <u>Journal</u> of Verbal Learning and Verbal Behavior, 1974, 13, 638-643.
- Rowe, E.J., & Paivio, A. Word frequency and imgery effects in verbal discrimination learning. <u>Journal of Experimental Psychology</u>, 1971, <u>88</u>, 319-326.
- Rowe, E.J., & Rose, R.J. Effects of orienting task, spacing of repetitions and list context on judgments of frequency. <u>Memory</u> and Cognition, 1977, 5, 505-512.
- Shepard, R.N. Recognition memory for words, sentences and pictures. Journal of Verbal Learning and Verbal Behavior, 1967, <u>6</u>, 156-163.
- Tulving, E. Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.) <u>Organization of Memory</u>. New York: Academic Press, 1972.
  Tversky, A., & Kahneman, D. Availability: A heuristic for judging frequency and probability. <u>Cognitive Psychology</u>, 1973, <u>5</u>, 207-232.
- Underwood, B.J. The attributes of memory. <u>Psychological Review</u>, 1969, <u>76</u>, 559-573.
- Underwood, B.J. Word recognition, memory and frequency information. Journal of Experimental Psychology, 1972, 94, 276-283.
- Underwood, B.J., & Freund, J.S. Relative frequency judgments and verbal discrimination learning. Journal of Experimental Psychology, 1970, 83, 279-285.
- Underwood, B.J., Zimmerman, J., & Freund, J.S. Retention of frequency information with observations on recognition and recall. Journal of Experimental Psychology, 1971, 87, 149-162.

Wallace, W.P., Murphy, M.D., & Sawyer, T.J. Imagery and frequency in verbal discrimination learning. <u>Journal of Experimental</u> <u>Psychology</u>, 1973, <u>101</u>, 201-219.

Ward, W.C., & Jenkins, H.M. The display of information and the judgment of contingency. <u>Canadian Journal of Psychology</u>, 1965, <u>19</u>, 231-241.

Watkins, M.J., & Tulving, E. Episodic memory: When recognition fails. <u>Journal of Experimental Psychology: General</u>, 1975, <u>104</u>, 5-29.
Williams, R.F., & Underwood, B.J. Encoding variability: Tests of the Martin hypothesis. <u>Journal of Experimental Psychology</u>, 1970, 86, 317-324.